



Editora
UFPel

Problemas Filosóficos

Uma Introdução à Filosofia

Rodrigo Cid
Luiz Helvécio Marques Segundo
(Organizadores)

DISSERTATIO
FILOSOFIA

**PROBLEMAS FILOSÓFICOS
UMA INTRODUÇÃO À FILOSOFIA**

Série Dissertatio Filosofia

**PROBLEMAS FILOSÓFICOS
UMA INTRODUÇÃO À FILOSOFIA**

Rodrigo Cid
Luiz Helvécio Marques Segundo
(Organizadores)

DISSERTATIO
FILOSOFIA
Pelotas, 2020

REITORIA

Reitor: Pedro Rodrigues Curi Hallal

Vice-Reitor: Luís Isaias Centeno do Amaral

Chefe de Gabinete: Taís Ullrich Fonseca

Pró-Reitor de Graduação: Maria de Fátima Cóssio

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

Pró-Reitor de Extensão e Cultura: Francisca Ferreira Michelin

Pró-Reitor de Planejamento e Desenvolvimento: Otávio Martins Peres

Pró-Reitor Administrativo: Ricardo Hartlebem Peter

Pró-Reitor de Infraestrutura: Julio Carlos Balzano de Mattos

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

Pró-Reitor de Gestão Pessoas: Sérgio Batista Christino

CONSELHO EDITORIAL DA EDITORA DA UFPEL

Presidente do Conselho Editorial: João Luis Pereira Ourique

Representantes das Ciências Agrônômicas: Guilherme Albuquerque de Oliveira Cavalcanti

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

Representantes da Área das Ciências Biológicas: Marla Piumbini Rocha

Representante da Área das Engenharias e Computação: Darci Alberto Gatto

Representantes da Área das Ciências da Saúde: Claiton Leoneti Lencina

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

Representante da Área das Ciências Humanas: Charles Pereira Pennaforte

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

EDITORIA DA UFPEL

Chefia: João Luis Pereira Ourique (Editor-chefe)

Seção de Pré-produção: Isabel Cochrane (Administrativo)

Seção de Produção: Gustavo Andrade (Administrativo)

Anelise Heidrich (Revisão)

Ingrid Fabiola Gonçalves (Diagramação)

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

Morgana Riva (Assessoria)

CONSELHO EDITORIAL

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

COMISSÃO TÉCNICA (EDITORIAÇÃO)

Prof. Dr. Juliano Santos do Carmo (Diagramador)
Prof. Dr. Juliano Santos do Carmo (Capista)
Profa. Luana Francine Nyland (Assessoria)

DIREÇÃO DO IFISP

Prof. Dr. João Hobuss

CHEFE DO DEPARTAMENTO DE FILOSOFIA

Prof. Dr. Juliano Santos do Carmo

Série Dissertatio Filosofia

A Série Dissertatio Filosofia, uma iniciativa do Núcleo de Ensino e Pesquisa em Filosofia (sob o selo editorial NEPFIL online) em parceria com a Editora da Universidade Federal de Pelotas, tem por objetivo precípuo a publicação de estudos filosóficos relevantes que possam contribuir para o desenvolvimento da Filosofia no Brasil nas mais diversas áreas de investigação. Todo o acervo é disponibilizado para download gratuitamente. Conheça alguns de nossos mais recentes lançamentos.

Estudos Sobre Tomás de Aquino

Luis Alberto De Boni

Do Romantismo a Nietzsche: Rupturas e Transformações na Filosofia do Século XIX

Clademir Luís Araldi

Didática e o Ensino de Filosofia

Tatielle Souza da Silva

Michel Foucault: As Palavras e as Coisas

Kelin Valeirão e Sônia Schio (Orgs.)

Sobre Normatividade e Racionalidade Prática

Juliano do Carmo e João Hobuss (Orgs.)

A Companion to Naturalism

Juliano do Carmo (Organizador)

Ciência Empírica e Justificação

Rejane Xavier

A Filosofia Política na Idade Média

Sérgio Ricardo Strefling

Pensamento e Objeto: A Conexão entre Linguagem e Realidade

Breno Hax

Agência, Deliberação e Motivação

Evandro Barbosa e João Hobuss (Organizadores)

Acesse o acervo completo em:

wp.ufpel.edu.br/nepfil

© **Série Dissertatio de Filosofia, 2020**

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

NEPFil online

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

Os direitos autorais estão de acordo com a Política Editorial do NEPFil online. As revisões ortográficas e gramaticais foram realizadas pelos autores e organizadores.

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

Dados Internacionais de Catalogação

N123 Problemas filosóficos: uma introdução à filosofia.
[recurso eletrônico] Organizadores: Rodrigo Cid; Luiz Helvécio Marques
Segundo – Pelotas: NEPFIL Online, 2020.
649p. - (Série Dissertatio Filosofia).
Modo de acesso: Internet
<wp.ufpel.edu.br/nepfil>
ISBN: 978-65-86440-43-0

1. Filosofia. 2. Problemas Filosóficos. I. Segundo, Luiz Helvécio Marques.
II. Cid, Rodrigo.

COD 100



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

SUMÁRIO

Introdução	16
Rodrigo Reis Lastra Cid Luiz Helvécio Marques Segundo	
1. Lógica	21
Éderson Safra Melo Marcio Kléos Freire Pereira	
2. Metafísica	79
Rodrigo Alexandre de Figueiredo	
3. Epistemologia	111
Delvair Custódio Moreira	
4. Ética	152
Bruno Aislã Gonçalves dos Santos Rafael Martins	
5. Estética	252
André Luiz Alves Pereira	
6. Filosofia da Ciência	306
Tiago Luís Teixeira de Oliveira	
7. Filosofia da Física	363
Diana Taschetto Thales Borrely	
8. Filosofia da Biologia	420
Sérgio Farias de Souza Filho	
9. Filosofia da Linguagem	453
Sajid Salles	

10. Filosofia da Mente	490
Samuel C. Bellini-Leite	
11. Filosofia Política	520
Everton Miguel Puhl Maciel	
12. Filosofia da Economia	552
Ramiro de Ávila Peres	
13. Filosofia da Religião	593
Luiz Helvécio Marques Segundo	

ORGANIZADORES

Luiz Helvécio Marques Segundo: Doutor em filosofia pelo PPG-Fil/UFSC. Atualmente é pesquisador vinculado à Associação Brasileira de Filosofia da Religião (ABFR). Tem atuado na interseção entre epistemologia, filosofia da religião e filosofia da ciência, investigando questões relativas ao debate entre ciência e religião, com particular ênfase à teoria da evolução biológica. Interessa-se também por abordagens evolutivas à moralidade e à cultura. Foi professor no Departamento de Filosofia da Universidade Federal de Ouro Preto entre 2016 e 2019.

Rodrigo Reis Lastra Cid: Professor Adjunto de Filosofia na Universidade Federal do Amapá. Editor Chefe do periódico *Investigação Filosófica*. Residência Pós-Doutoral em Filosofia realizada na Universidade Federal de Minas Gerais. Doutor e Mestre em Lógica e Metafísica (Filosofia) pelo Programa de Pós-Graduação em Lógica e Metafísica da Universidade Federal do Rio de Janeiro, com período sanduíche na Université Catholique de Louvain. Bacharel em Filosofia pela Universidade Federal em Ouro Preto. Ex-Professor de Filosofia na Universidade Federal do Rio de Janeiro, no Instituto Federal de Minas Gerais e na Faculdade Dom Luciano Mendes. Pesquisador na área da Metafísica. Membro do GT de Metafísica Analítica da ANPOF, líder do Grupo de Pesquisa *Investigação Filosófica* (DGP/CNPq) e membro da Society for the Metaphysics of Science.

AUTORES

André Luiz Alves Pereira: Professor de História na SME de Florianópolis. Graduado em História pela Universidade Federal de Santa Catarina; Mestre em História pela Universidade Federal de Santa Catarina. Atua nas áreas: História Social do Brasil Monárquico e História Social do Brasil Republicano, com ênfase nas relações entre História e Literatura. Investiga paralelamente questões sobre Filosofia das Ciências Sociais e sobre o valor cognitivo da literatura.

Bruno Aislã Gonçalves dos Santos: graduado em Filosofia (Bacharelado e

Licenciatura) pela Universidade Federal de Ouro Preto (UFOP), é Mestre e Doutor em Ética e Filosofia Política pela Universidade Federal de Santa Catarina (UFSC) e fez estágio de doutoramento na University of St. Andrews (Escócia) sob a orientação de Tim Mulgan. Estuda o utilitarismo clássico e contemporâneo voltado as discussões sobre a justiça distributiva e justiça global. Tem experiência na área de Filosofia, com ênfase em Ética e Filosofia Política. Atua principalmente nas seguintes áreas: Ética, Filosofia Política, Filosofia do Direito, Filosofia da Linguagem e Filosofia da Educação.

Delvair Custodio Moreira: Professor adjunto na Universidade Federal do Maranhão. Doutor em Lógica e Epistemologia pela Universidade Federal de Santa Catarina (2017); mestre em Lógica e Epistemologia pela Universidade Federal de Santa Catarina (2013); e bacharel em Filosofia pela Universidade Federal de Ouro Preto (2010). Tem experiência na área de filosofia com ênfase em epistemologia e epistemologia social. Seus principais temas de interesse são: epistemologia do testemunho, conhecimento coletivo e epistemologia da religião.

Diana Taschetto: Doutoranda em Filosofia da Física pela Universidade de São Paulo (USP); mestre em Filosofia pela Universidade de São Paulo (USP), graduanda em Física bacharelado pela Universidade de São Paulo (USP), graduada em Filosofia pela Universidade do Vale do Rio dos Sinos (UNISINOS). Tem experiência em filosofia da física, filosofia da ciência e epistemologia, tendo investigado, enquanto graduanda, questões relacionadas ao falseacionismo (Popper), ao anarquismo epistemológico (Feyerabend) e à metodologia dos programas de pesquisa científicos (Lakatos). Investigou a estratégia de justificação coerentista na epistemologia moral de John Rawls, suas raízes científicas e sua relação com a filosofia de Quine em seu último ano na graduação da UNISINOS. Enquanto aluna de mestrado em Filosofia da Física, na Universidade de São Paulo investigou controvérsias filosóficas nos fundamentos da teoria das cordas. Atualmente investiga observáveis em física quântica.

Ederson Safra Melo: Professor no Departamento de Filosofia da Universidade Federal do Maranhão. Fez graduação em Filosofia pela Universidade Estadual de Londrina (2009); concluiu o mestrado (2012 com bolsa CNPq) e o doutorado (2017, com bolsa CAPES) na área de Lógica e Epistemologia pelo Programa de

Pós-Graduação em Filosofia da Universidade Federal de Santa Catarina. Durante o doutorado foi pesquisador visitante na Brown University, EUA, (2015 - 2016, com bolsa CAPES/PDSE). Tem experiência nas áreas Filosofia da Lógica, Lógica, Epistemologia e Filosofia da Linguagem. Atua principalmente nos seguintes temas: Paradoxos Semânticos, Dialeteísmo, Teorias da Negação e Teorias da Verdade. É líder do Grupo em Filosofia da Lógica e da Linguagem (GFILL/CNPq) e membro do Grupo de Estudos em Lógica e Filosofia Formal (GELF/CNPq), atuando como coordenador de pesquisa em Filosofia da Lógica. É coordenador do Programa de Pós-Graduação em Filosofia da UFMA.

Everton Miguel Puhl Maciel: professor de Filosofia Universidade Federal do Amapá, bacharel em Filosofia pela Universidade Federal de Pelotas, mestrado pela mesma instituição na linha de pesquisa Ética e Filosofia Política; doutor em Filosofia pela Pontifícia Universidade Católica do Rio Grande do Sul, com estágio doutoral em Birkbeck, Universidade de Londres.

Marcio Kléos Freire Pereira: Professor no Departamento de Filosofia da Universidade Federal do Maranhão, professor permanente no PROF-FILO (Mestrado Profissional em Filosofia – núcleo UFMA) e no PPGFIL/UFMA (Mestrado Acadêmico em Filosofia). Também é líder do GELF/UFMA (Grupo de Estudos em Lógica e Filosofia Formal) e pesquisador do GFILL/UFMA (Grupo em Filosofia da Lógica e da Linguagem). Concluiu bacharelado (1993), licenciatura (1995) e mestrado (1996) em Filosofia pela Universidade Federal da Paraíba, e doutorado (2015) em Filosofia pela Universidade Federal de Santa Catarina. Tem experiência nas áreas de Lógica e Epistemologia, atuando principalmente nos seguintes temas: lógicas modais (lógica do anúncio público, lógica epistêmica dinâmica), filosofia analítica (atitudes proposicionais, teoria do conhecimento), teoria da argumentação (lógica informal) e sua relação com o ensino de Filosofia.

Ramiro de Ávila Peres: Analista do Banco Central. Graduação em direito, mestrado e doutorado em filosofia pela Universidade federal do Rio Grande do Sul. Área: Filosofia, Política e Economia; Ética e Finanças.

Rafael Martins: Doutorando na University of Kansas, onde é também professor assistente. Suas pesquisas concentram-se nas áreas da filosofia política, ética

normativa, metaética, filosofia do direito, razão prática e história do pensamento econômico.

Rodrigo Figueiredo: Possui graduação em filosofia pela Universidade Federal de Ouro Preto (2009), mestrado em Lógica e Metafísica pela Universidade Federal do Rio de Janeiro (2012) e doutorado em Lógica e Metafísica pela Universidade Federal do Rio de Janeiro (2017). Atualmente é professor da Faculdade Dom Luciano Mendes. Atua principalmente nos seguintes temas: propriedades, universais, compromisso ontológico, particulares e necessidade. Foi Bolsista da CAPES no Programa de Doutorado Sanduíche no Exterior (PDSE).

Sagid Salles: Graduado (licenciatura e bacharelado) em filosofia pela Universidade Federal de Ouro Preto. Mestre pela Universidade Federal do Rio de Janeiro, pelo Programa de Pós-Graduação Lógica e Metafísica (PPGLM). Doutorado com louvor também pelo PPGLM, com período de sanduíche realizado na Universidade de Miami.

Samuel Bellini-Leite: Doutorado em Filosofia - Lógica, Ciência, Mente e Linguagem - na Universidade Federal de Minas Gerais (UFMG). Mestrado em Filosofia - Filosofia da Mente, Ciência Cognitiva e Semiótica - pela Universidade Estadual Paulista (UNESP). Graduado em Psicologia pelo Centro de Ensino Superior de Juiz de Fora. Se concentra nas áreas de Psicologia Cognitiva, Filosofia da Psicologia, Psicologia da Educação, Filosofia da Mente e Ciência Cognitiva Teórica. Atualmente leciona na Universidade do Estado de Minas Gerais.

Sérgio Farias de Souza Filho: Pesquisador de Pós-Doutorado da Universidade Federal do Rio de Janeiro e do Conselho Nacional de Desenvolvimento Científico e Tecnológico. Doutorado em Filosofia, King's College London. Mestrado em Filosofia pelo Programa de Pós-Graduação Lógica e Metafísica da Universidade Federal do Rio de Janeiro. Graduação em Filosofia pela Universidade Federal de Pernambuco. Áreas: Filosofia da Mente, Filosofia da Linguagem, Metafísica e Filosofia da Biologia.

Thales Borrely dos Santos: Bacharel em física pelo Instituto de Física da

Universidade de São Paulo (IFUSP). Foi aluno de iniciação científica do laboratório de cristais iônicos, filmes finos e datação do IFUSP. Nesse período foi bolsista do CNPq com quatro projetos distintos, todos na área de filmes finos de óxidos produzidos por deposição assistida por feixe de íons. Entre 2014 e 2017 fez mestrado acadêmico nesse mesmo laboratório. Ainda em 2017 ingressou no programa de doutorado onde desenvolve o projeto "Células fotovoltaicas baseadas em pontos quânticos de submonocamada de InAs" sob orientação do professor Dr. Alain André Quivy.

Tiago Luís Teixeira de Oliveira: Professor do departamento de filosofia do Colégio Pedro II. Doutor em filosofia pela Universidade Federal de Minas Gerais com estágio pós-doutoral na Universidade do Estado do Rio de Janeiro. Áreas: Epistemologia, Filosofia da Ciência e Ensino de Filosofia.

Introdução

Em primeiro lugar, gostaríamos de agradecer aos autores, que se dedicaram com afinco ao projeto, ao Grupo de Pesquisa Investigação Filosófica (DGP/CNPq), do qual alguns de nós somos parte, e ao Prof. Dr. Juliano do Carmo, ao NEPFIL online e à Editora da UFPel, por todo trabalho envolvido na edição e publicação deste livro. O objetivo deste livro é ser uma Introdução à Filosofia para leigos e para estudantes de primeiro ou segundo período de Filosofia e de outras áreas que contam com essa disciplina. Você vai notar que este livro possui uma organização levemente diferente da maioria dos livros introdutórios da disciplina. Em vez de uma divisão por períodos históricos da filosofia, ou por autores, ou por temas gerais, decidimos fazer uma partição em subáreas da filosofia e seus problemas. Assim, cada um dos capítulos apresentará uma subárea da filosofia com seus maiores problemas e teorias que tentam solucioná-los, tal como algumas questões de revisão, outras de discussão e uma sugestão de bibliografia para o estudante iniciante ou para o leitor interessado.

Isso é importante, por apresentar a filosofia como um conjunto de problemas em aberto, sobre os quais o leitor estará convidado a refletir. A visão por trás desse livro traz consigo uma certa noção de filosofia, que a vê como a tentativa de solucionar certos problemas que aparecem na teorização filosófica. Esses problemas têm a característica de não poderem ser resolvidos por experimentos ou por cálculos. Eles são problemas intrinsecamente argumentativos, i.e., apenas solucionáveis (se é que são solucionáveis) por argumentação.

A filosofia tem muitas áreas e podemos explicar, pedagogicamente, como podemos pensar sua taxonomia. Há muitas formas diferentes de se traçar uma divisão da filosofia. Vamos apresentar uma delas, que promete ajudar o estudante a compreender a área. Costumamos dividir a filosofia em três grandes áreas propriamente filosóficas e uma área técnica, que seriam, respectivamente, a metafísica, a epistemologia, as teorias do valor e a lógica.

A lógica é importante, pois, como somos uma área que trabalha com problemas argumentativos, temos de entender o funcionamento da argumentação ou do raciocínio, para não utilizarmos falácias ou errarmos na

consistência da nossa construção teórica. A lógica clássica tem a propriedade de nos ajudar a compreender quando estamos construindo um argumento válido e quando não estamos. Sem essa área técnica fundamental, não temos prosseguir, com rigor, em nossas investigações. Sem ela, não conseguiríamos avaliar as nossas teorias. A ideia da lógica é que, se começarmos com hipóteses que sabemos serem verdadeiras e utilizarmos lógica para derivar conclusões, podemos estar certos de que as conclusões serão também verdadeiras. Mas saber que as hipóteses iniciais são verdadeiras, isso não temos como saber *somente* com base na lógica; isso vai exigir uma discussão muito mais aprofundada, levando em conta os resultados científicos, nossas melhores intuições intelectuais, resultados de outras áreas da filosofia etc. É claro que a área da lógica é muito maior do que a lógica clássica e serve para muitas outras coisas do que meramente lidar com o raciocínio e com a argumentação válida. Há também muitas outras lógicas diferentes da lógica clássica, mas como essa pretende ser uma introdução à filosofia, optamos por apresentar apenas a lógica clássica como uma preparação para a argumentação rigorosa, deixando o conhecimento mais sistemático de lógicas não clássicas para um outro livro.

Com relação às áreas de maior conteúdo propriamente filosófico, temos a metafísica, a epistemologia e as teorias do valor, conforme mencionamos. A metafísica trata daquilo que existe na realidade. As questões dessa área envolvem se perguntar se existem certas entidades e quais as categorias mais gerais em que essas entidades se organizam. Por exemplo, questões intimamente metafísicas seriam aquelas sobre a existência de Deus, a existência de livre-arbítrio, a existência de espaço, de tempo, de matéria e sobre a categorização geral dessas entidades. Como devemos pensar sua divisão mais geral? O mundo se divide em objetos particulares que têm certas propriedades? São o objeto individual e as propriedades as categorias mais gerais em que todas as entidades se dividem? Por exemplo, devemos pensar um pássaro preto como algo particular que tem as propriedades de ser pássaro e de ser preto? O que é esse *algo particular*? O que são essas *propriedades*? As respostas para essas perguntas, só obteremos depois de muita discussão metafísica.

Porém não apenas de metafísica vive o filósofo, mas também de epistemologia, que é a área em que se discute se e como obtemos conhecimento sobre as coisas que existem no mundo (e talvez até sobre as

coisas que não existem). Podemos falar das coisas que existem, mas como saber se sabemos sobre o que estamos falando ou se apenas achamos que sabemos? Do que se constitui o conhecimento? É ele crença verdadeira e justificada? Quais são as condições para dizermos que uma pessoa tem conhecimento de algo? O conhecimento e suas justificações são algo de individual ou de social? Desde o problema do ceticismo e da natureza do conhecimento até o problema da qualidade e da natureza dos nossos métodos e justificações são tratados pela epistemologia. Essa área promete-nos dizer o que e como sabemos e o que e como podemos saber.

Por sua vez, as teorias do valor não se perguntam propriamente o que existe e nem como obtemos conhecimento do que existe, mas elas apreendem a realidade em seu aspecto avaliativo, valorativo. Nelas, queremos mais saber como devem ser as coisas do que propriamente como as coisas são. Queremos saber o que é certo ou errado, bom ou mau, belo ou feio etc. Nessa área, trabalhamos pelo menos com dois valores fundamentais: o moral/ético e o estético. Como os nomes deixam transparecer, a ética ou teoria da moral é uma das teorias do valor, que lida com os valores morais, do certo e do errado e do bom e do mau. E a estética é aquela área que lida com o valor estético, perguntando-se o que é feio ou belo, ou, de modo geral, o que é estético ou não estético. Essas áreas do valor são compostas também de metafísica e epistemologia do próprio valor, quando adentramos na área conhecida pelo nome de metaética ou metaestética. De modo geral, essas áreas do valor se dividem em três partes pelo menos: (1) a estética prática e ética prática, nas quais nos perguntamos se uma certa coisa ou ação é ela mesma valorosa (feia/bonita, boa/má, certa/errada etc); (2) a ética e a estética normativas, nas quais nos perguntamos qual é a definição adequada desses valores, i.e., o que é o bem, o mal, o correto, o incorreto, o feio, o belo, o estético e o não estético; e (3) a metaética e a metaestética, nas quais nos perguntamos sobre a existência e a conhecibilidade desses valores éticos e estéticos, misturando metafísica e epistemologia às teorias do valor.

Além dessas disciplinas básicas da filosofia, as grandes áreas, temos também áreas mistas, de problemas filosóficos (metafísicos, epistêmicos, valorativos e lógicos) com relação a cada uma das disciplinas acadêmicas ou delas em geral, como a filosofia da ciência, a filosofia da física, a filosofia da biologia, a filosofia da política, a filosofia da economia, a filosofia da mente, a

filosofia da linguagem, a filosofia da religião, dentre outras (tal como a filosofia da lógica, que não é meramente uma área técnica como a lógica, mas genuinamente filosófica, tal como as outras filosofias de cada disciplina). Essas áreas são um misto de metafísica, epistemologia, teorias do valor e questões de consistência com relação às áreas de onde provêm essas disciplinas, como a ciência em geral, a física, a biologia, a política, a economia, a ciência cognitiva e as questões sobre a mente, a linguística e as questões sobre a linguagem, a teologia e as questões sobre a religião etc.

Isso significa que todas as áreas de investigação têm alguns problemas que só poderiam ser solucionados por argumentação, e não por experimentos ou cálculos. Podemos nos perguntar sobre as propriedades de um fóton na física, mas quem responde se há tais coisas como fótons ou se eles são meros instrumentos teóricos para a previsão, não existindo verdadeiramente, isso é a filosofia da física. Podemos nos perguntar qual a probabilidade de termos um filho de olho azul para o biólogo, mas saber qual a natureza da vida, isso já é um assunto da filosofia da biologia. Podemos compreender as ações dos Estados, na política internacional e na economia, como fazendo parte de um grupo político ou mesmo como inseridas em uma certa política econômica, mas compreender qual a natureza do Estado e de sua justificação e quais são os valores fundamentais em política e economia, isso já é parte da filosofia da política e da filosofia da economia. Podemos observar um cérebro em uma máquina de ressonância magnética com as técnicas experimentais da psicologia e da medicina, mas só podemos saber qual a natureza da mente, se idêntica ou distinta do cérebro, por meio das discussões em filosofia da mente. Embora podemos saber empiricamente como os indivíduos aprendem uma linguagem, quando estudamos linguística, saber qual a natureza da linguagem, da significação e da referência, isso só podemos entender nas discussões de filosofia da linguagem. Da mesma forma, podemos nos perguntar inúmeras questões de teologia, que conectam história, sociologia, entre outras disciplinas acadêmicas, mas saber qual a natureza da religião, se Deus existe e se e como podemos saber tais coisas, isso já está no domínio da filosofia da religião.

De um modo geral, queríamos mostrar que a filosofia tem suas próprias áreas, mas tem também subáreas em interdisciplinaridade com as ciências. As ciências e as disciplinas acadêmicas em geral têm problemas, cuja a solução pode ser encontrada empiricamente, por meio de experimentos,

entrevistas, documentos, ou formalmente, por meio de cálculos etc, porém os problemas das filosofias dessas disciplinas são justamente os problemas mais fundamentais dessas disciplinas, que fundam o quadro conceitual e de pesquisa das mesmas, e que só poderiam ser respondidos por meio de uma extensa argumentação. A filosofia de cada uma dessas disciplinas é uma mistura (a) de metafísica da disciplina, perguntando-se qual a natureza e justificação das entidades que formam o quadro teórico da disciplina, i.e., quais delas realmente existem na realidade (não sendo *meros* instrumentos para obter resultados) e quais as suas categorizações fundamentais, naturezas e funções teóricas, (b) de epistemologia dessa disciplina, ao nos perguntarmos sobre como tal área poderia obter conhecimento de modo confiável e se ela de fato o faz, e (c) de teorias do valor, ao nos perguntarmos, por exemplo, sobre qual o valor moral do desenvolvimento de uma certa disciplina e, finalmente, (d) a lógica, que perpassa tudo, ao nos perguntarmos sobre a consistência das teses assumidas pela disciplina.

Não é à toa que falamos que a filosofia é mãe de todas as ciências. Ela é mãe tanto num sentido geracional, pois a maioria das ciências foram desenvolvimentos particulares, empíricos ou de cálculo, das questões filosóficas sobre um certo tema, quanto num sentido fundamental, pois é a filosofia que fornece os fundamentos mais gerais, metafísicos, epistêmicos, valorativos e lógicos, na construção do quadro teórico de cada ciência e na qualidade dos métodos para se aproximar da verdade que a disciplina utiliza. A biologia, por exemplo, foi criada por Aristóteles, tal como os primeiros desenvolvimentos da meteorologia (na verdade, esta começa com os pré-socráticos), e grande parte do fundamento da física atual jaz nas discussões de filosofia da física.

Sendo você filósofo ou cientista de outra área, a filosofia promete te dar as bases para uma reflexão geral ou sobre sua área, que seja muito mais informada e profunda do que aquilo que você poderia realizar sem ela.

Rodrigo Reis Lastra Cid
Luiz Helvécio Marques Segundo
(Organizadores)

1

Lógica

Éderson Safra Melo
Marcio Kléos Freire Pereira

1. Introdução: lógica e inferência

A palavra ‘lógica’ admite várias acepções no português. Podemos nos deparar com expressões como “a lógica do mercado de ações”, “a lógica da vida”, dentre outras, com acepções como *o funcionamento* ou *o sentido de algo*. Todavia, a palavra “lógica” serve também para se referir à disciplina Lógica, que teve sua origem na Antiguidade, com os trabalhos de Aristóteles, e tem se desenvolvido e se prestado a aplicações em diversas áreas. Aristóteles levou o crédito de ser conhecido como “o pai da Lógica”, por ser considerado o primeiro a estabelecer um sistema de lógica, conhecido como *teoria do silogismo*. Embora tenha existido inicialmente uma tradição paralela independente de Aristóteles (a lógica estoica), por razões históricas, a Lógica se resumiu basicamente à lógica aristotélica até meados do século XIX, quando foi revolucionada pelos trabalhos de George Boole (1815-1864) e Gottlob Frege (1848-1925). A publicação da obra *Conceitografia* (1879), de Frege, representa um marco na história dessa disciplina, fazendo com que passasse a ser qualificada como *simbólica* ou *matemática*, devido ao uso de notações simbólicas similares às da Matemática, além de novas aplicações voltadas para esta ciência. Desde o século XX, a Lógica tem permitido diversas aplicações e desenvolvimentos em numerosas áreas, tanto dentro da Filosofia como em algumas ciências – por exemplo, na Linguística, na Ciência da Computação, na Inteligência Artificial, entre outras.¹

¹Para história da Lógica, veja KNEALE & KNEALE, 1980.

Aqui não trataremos dessas aplicações e nos restringiremos à análise de argumentos, considerada, às vezes, como a aplicação *canônica* da Lógica (cf. PRIEST, 2014). Nessa acepção, a Lógica explora os mecanismos ou métodos de inferências que podem ser utilizados para análise de argumentos. Nessa perspectiva, podemos adotar a seguinte noção de Lógica: “LÓGICA é a ciência que estuda princípios e métodos de inferência, tendo o objetivo principal de determinar em que condições certas coisas se seguem (são consequências), ou não, de outras” (MORTARI, 2016, p. 14).

Como ponto de partida, tomamos a Lógica (com a inicial maiúscula para denotar a disciplina intelectual) como um conjunto de lógicas (com a inicial minúscula) que constituem teorias de inferências. Mas o que é uma inferência?

Para ilustrar o conceito de inferência, vejamos rapidamente uma anedota. Três lógicos entram em um bar e o garçom pergunta: *todostrês* vão tomar cerveja? Suponha que nenhum dos lógicos sabe das intenções dos demais. O primeiro lógico diz que não sabe se os três pretendem beber cerveja. Depois, o segundo lógico também diz que não sabe a resposta. Por fim, o terceiro lógico convictamente diz: sim, nós *todos* iremos tomar cerveja. A questão que se coloca é: por que o terceiro lógico tem a resposta para a pergunta do garçom e os demais não? A resposta é que, no contexto da anedota, o terceiro lógico pode inferir a informação que responde à pergunta a partir das respostas dos outros dos lógicos – i.e., foi possível extrair a informação que serviu como resposta para o garçom. Mas como isso foi possível? (O leitor pode pensar um pouquinho para tentar descobrir como isso se deu antes de ler o próximo parágrafo).

A pista para saber como o terceiro lógico pôde responder à pergunta de maneira convicta está na expressão “*todos três*” contida na pergunta do garçom (“*todostrês* vão tomar cerveja?”). Se o primeiro lógico que respondeu à pergunta não pretendesse tomar cerveja, a resposta dele seria “não”, deixando claro que nem todos iriam tomar cerveja (já que ele mesmo não iria tomar cerveja). Como ele pretendia tomar cerveja, mas não pode falar sobre as pretensões de seus colegas lógicos, apenas diz: “eu não sei”. O mesmo raciocínio vale para o segundo lógico que respondeu da mesma maneira, com a diferença de que o segundo lógico já havia concluído que o primeiro também iria tomar cerveja. Como o segundo lógico ainda não tinha a resposta do terceiro lógico, ele apenas respondeu dizendo que não sabia. Agora, sabendo das informações contidas nas respostas dos dois primeiros, o terceiro lógico teve condições de inferir que os

dois iriam tomar cerveja, já que os dois primeiros responderam que não sabiam se todos os três iriam tomar cerveja. Com base nessas informações e sabendo que ele próprio irá tomar cerveja, ele conseguiu *inferir* que todos os três lógicos irão tomar cerveja.²

O processo de extrair informações a partir de outras é o que entendemos por processo de inferência. Assim, podemos inferir ou extrair consequências daquilo de que dispomos como informação (algo em que acreditamos ou que estamos supondo ser verdadeiro). Na anedota acima, com base nas informações disponíveis, o terceiro lógico conseguiu obter como consequência algo que ele não sabia antes (i.e, que todos três iriam tomar cerveja). De modo geral, a Lógica estuda princípios de inferências que autorizam ou não certas informações serem consequências de outras. Grosso modo, a Lógica diz respeito à relação de consequência estabelecida nas diferentes lógicas (como na lógica clássica, por exemplo).

A presente exposição não pretende desempenhar o papel de um curso básico de Lógica que instrumentaliza o leitor com todos os conceitos, técnicas e métodos fundamentais daquela disciplina. Há alguns manuais introdutórios que cumprem muito bem esse papel. Indicaremos alguns desses manuais ao longo do texto. Aqui, pretendemos apenas oferecer uma visão bastante geral sobre a Lógica e como ela é entendida em sua aplicação canônica, a análise de argumentos. Para tanto, na próxima seção, trataremos de argumentos bons e ruins do ponto de vista lógico. Depois, na seção 3, faremos uma apresentação sucinta de uma lógica muito simples: o Cálculo Proposicional Clássico (CPC). Especificamente, apresentaremos uma linguagem para o CPC (na subseção 3.1), depois veremos sua parte semântica (subseção 3.2) e, por fim, apresentaremos sua contraparte sintática (na subseção 3.3). Encerraremos na seção 4, fazendo algumas considerações breves sobre a relação entre as contrapartes sintática e semântica do CPC, bem como mencionando algumas lógicas não clássicas e alguns problemas filosóficos que surgem na Lógica.

²A anedota acima pode ser encontrada no documentário sobre Lógica produzido pela BBC intitulado *The Joy of Logic*. O documentário com legendas em português pode ser encontrado através do seguinte link:<https://www.youtube.com/watch?v=dr1PO-AOeFY>.

2. Argumentos

2.1. O que é um argumento?

Saber argumentar bem é uma habilidade importante em numerosas atividades cotidianas, desde aquelas mais informais – por exemplo, tentar convencer um amigo acerca de determinada informação – até contextos mais sérios e de grande impacto social – por exemplo, quando um cientista precisa convencer seus pares acerca de algum resultado, ou quando um parlamentar precisa justificar um projeto de lei. Embora o convencimento possa eventualmente depender também de elementos subjetivos – por exemplo, a eloquência ou o carisma pessoal do argumentador, ou, ainda, as vantagens e desvantagens pessoais em se concordar com sua argumentação –, trataremos aqui apenas do convencimento estritamente racional, deixando os demais fatores para outros campos de estudo (como a Retórica).

Na Filosofia, a argumentação é fundamental, visto que somos cobrados acerca das razões para se aceitar as teses que sustentamos. Frequentemente, precisamos também argumentar contra as teses que criticamos ou não achamos razoáveis por algum motivo. Como poderíamos trabalhar com qualquer problema filosófico sem recorrer a argumentos? Poderíamos fundamentar, por exemplo, alguma tese sobre a confiabilidade das teorias científicas ou sobre direitos humanos fundamentais com base em evidências observáveis? Não podemos decidir sobre a maioria das questões filosóficas a partir de dados empíricos; restando-nos, assim, recorrer à argumentação. Em síntese, na Filosofia se trabalha prioritariamente com argumentos e o sucesso de um trabalho filosófico depende necessariamente de se lidar bem com construções argumentativas. Bom, mas o que é um argumento?

Um **argumento** é um conjunto de sentenças no qual uma dessas sentenças (a *conclusão*) se distingue das demais (as *premissas*) por ser consequência delas. Dizendo de outra maneira, as premissas são evidências que garantiriam a conclusão.

A conclusão é a sentença que deve ser estabelecida ou sustentada pelo argumento. Podemos ter argumentos com apenas uma premissa até argumentos

com um número muito grande, porém sempre finito, de premissas. Ao caracterizar argumentos como um conjunto de sentenças, é importante destacar que não podemos ter o caso em que o referido conjunto de sentenças seja vazio, nem o caso em que o conjunto de sentenças seja infinito.³ No primeiro caso, não haveria nada para estabelecer algo (nenhuma premissa), nem nada a ser estabelecido (nenhuma conclusão) e, portanto, não teríamos um argumento. No segundo caso, sempre haveria razões adicionais sem que a conclusão pudesse ser estabelecida e, assim, nunca terminaríamos o argumento.

Feita essa breve caracterização, vamos conferir alguns exemplos de argumentos. Para facilitar a explicação de nossos exemplos, indicaremos a conclusão pela letra maiúscula em negrito '**C**' e as premissas pelas letras maiúsculas em negrito '**P₁, P₂, ...**', com índices, respectivamente, para enumerar a premissa 1, premissa 2, etc.

Consideremos, então, o seguinte conjunto de afirmações:

P₁: Se Deus é perfeito, então não está sujeito a carência alguma.

P₂: Deus é perfeito.

C: Logo, Deus não está sujeito a carência alguma.

De acordo com a definição dada acima, temos um argumento. Há um conjunto de afirmações (**P₁, P₂, C**), em que **P₁** e **P₂** têm o papel de sustentar a afirmação **C**.

Podemos extrair mais de uma conclusão nesse argumento, além da afirmação **C**? Estritamente, pensando apenas neste conjunto de afirmações, temos apenas a conclusão enunciada em **C**. No geral, dizemos que um argumento tem apenas uma conclusão, de modo que se há mais de uma conclusão, há mais de um argumento. Na prática, entretanto, é muito comum encadearmos argumentos, tirando conclusões intermediárias a fim de chegar na conclusão

³Na filosofia da lógica há uma discussão sobre qual entidade é um portador de verdade por excelência, em que se discute qual entidade (sentenças, proposições, enunciados) veicula ou porta primariamente a verdade. Neste texto não discutiremos a questão dos portadores de verdade e também não faremos um uso técnico dos termos 'sentenças', 'afirmações', 'proposições' e 'enunciados', tomando tais termos como intercambiáveis. Para uma discussão sobre portadores de verdade, ver HAACK, 2002, cap. 6.

desejada. Podemos usar a conclusão de um determinado argumento como premissa de outro argumento. Considere, por exemplo, o seguinte argumento:

P₁: O engano depende de alguma carência.

P₂: Deus não está sujeito a carência alguma.

C: Logo, Deus não é enganador.

Note que a conclusão **C** do primeiro argumento está sendo usada como a premissa **P₂** do segundo argumento. Ser premissa ou conclusão depende do papel que as informações desempenham nos argumentos.

Considerando a maneira como os argumentos foram organizados acima, fica fácil identificar o que é premissa e o que é conclusão. Todavia, na prática – nos livros de Filosofia, por exemplo – os argumentos não aparecem organizados esquematicamente com premissas e depois conclusão, e muito menos com símbolos (como **C**, **P₁**, **P₂**, ...). Por exemplo, o último argumento acima é um esquema simplificado de um argumento famoso, apresentado por René Descartes (1973, p. 120) na sua *Meditação Terceira* que adaptamos livremente para ilustrar essa flexibilidade na disposição de premissas e conclusão:

Deus [...] que possui todas as altas perfeições de que o nosso espírito pode ter alguma ideia, sem, no entanto, compreendê-las, que não está sujeito a carência alguma e que nada tem de todas as coisas que assinalam alguma imperfeição. Daí é bastante evidente que ele não pode ser enganador, posto que é manifesto pela luz natural que o engano depende de alguma carência.

Diferente do arranjo esquemático que fizemos com os argumentos apresentados anteriormente, note que, na apresentação de Descartes, a conclusão não aparece depois das premissas. No trecho citado, a conclusão aparece no meio do argumento, e não no final. A conclusão poderia aparecer também no começo do argumento, antes de todas as premissas. De fato, a ordem das afirmações não é determinante para a classificação do que é premissa ou conclusão nos argumentos. Na prática, as premissas e conclusões podem aparecer em

qualquer posição do argumento. Isso pode dificultar um pouco o nosso trabalho de identificar o que é premissa e o que é conclusão de um dado argumento.

Felizmente, há alguns indicadores que podem ajudar a determinar as expressões que desempenham o papel de premissa ou de conclusão. Por exemplo, no argumento de Descartes apresentado acima, temos expressões que indicam premissas e conclusões. Especificamente, o trecho “daí, é bastante evidente que” indica que a afirmação logo após o mesmo é a conclusão do argumento. Depois dessa conclusão, aparece a expressão “posto que”, prenunciando uma razão ou justificativa para sustentar a conclusão. Essas expressões são chamadas de marcadores ou indicadores de premissas ou conclusão. Nos quadros abaixo, há alguns exemplos de indicadores de premissas e de indicadores de conclusão.

Indicadores de premissa: pois, posto que, porque, já que, uma vez que, ora, desde que, em virtude de, sabemos que, a razão é que, etc.

Indicadores de conclusão: logo, portanto, conseqüentemente, segue-se que, por conseguinte, assim, então, por isso, de onde resulta que, temos que, etc.

Esses indicadores funcionam como uma espécie de guia no sentido de que nos ajudam a determinar o que é premissa e o que é conclusão dos argumentos. Porém, nem todos os argumentos exibem tais indicadores. Alguns argumentos contêm apenas indicadores de premissas, outros apenas indicadores de conclusão, e ainda há argumentos sem quaisquer indicadores explícitos. Contudo, mesmo sem indicadores, um argumento minimamente bem construído deve permitir a identificação do que é premissa e do que é conclusão recorrendo-se ao contexto do argumento (por exemplo, o autor de um texto pode ter deixado claro na introdução o que estaria assumindo ou o que pretende estabelecer em alguma parte do texto).

2.2. Validade e correção de argumentos

Como vimos, em um contexto lógico, um argumento tem a finalidade de estabelecer uma conclusão, sustentada através de suas premissas. Por vezes, um argumento pode ter premissas implícitas e até redundantes (premissas contendo informações repetidas com outras palavras). Se alguma de suas premissas não for aceita, o argumento pode não atingir sua finalidade – caso as demais premissas não garantam juntas a conclusão desejada. Além disso, é possível tentar fornecer razões para as premissas de um argumento recorrendo a um novo argumento (lembre das cadeias de argumentos mencionadas anteriormente).

Todavia, mesmo com premissas verdadeiras e aceitas pelo nosso interlocutor, há ainda a possibilidade de um argumento não ter êxito em sustentar a sua conclusão. A passagem das premissas para a conclusão pode não funcionar se a conclusão não for uma consequência das premissas.

Além disso, nem todos os argumentos sustentam sua conclusão com a mesma força. Na Lógica, consideramos como paradigmáticos os argumentos classificados como *válidos*. Para tornar clara a noção de validade, consideremos o seguinte argumento:

P₁: Ou João está na universidade, ou está em casa.

P₂: João não está em casa.

C: Por conseguinte, João está na universidade.

Podemos dizer que a conclusão está justificada pelas premissas? Aparentemente, sim. Note que não precisamos ter conhecimento sobre o contexto do argumento (não precisamos saber quem é João e nem em qual universidade ele estuda) para sabermos que a conclusão é uma consequência inevitável das premissas. Mesmo não sabendo nada a respeito de João, dizemos que se for verdade que ou ele está na universidade ou ele está em casa, e que, além disso, se for verdade que ele não está em casa, somos levados a concluir que é verdade que João está na universidade. Ou seja, não é possível que as premissas sejam verdadeiras e a conclusão seja falsa, ao mesmo tempo. Dizemos, nesses casos, que **C** é uma consequência lógica de **P₁** e **P₂**.

Uma conclusão é **consequência lógica** de um conjunto de premissas se não é possível ter uma situação em que todas aquelas premissas sejam verdadeiras e a conclusão em questão seja falsa.

Os argumentos cujas premissas e conclusões exibem essa relação são chamados de *válidos*. Assim, podemos apresentar a noção de argumento válido da seguinte maneira:

Um **argumento é válido** se e somente se sua conclusão for uma consequência lógica de suas premissas.

Diante dos últimos quadros de definições, temos que, em um *argumento válido*, se as premissas forem verdadeiras, a conclusão necessariamente será verdadeira. Observe que a noção de validade não exige que as premissas sejam de fato verdadeiras; a validade de um argumento demanda apenas que não pode haver uma situação possível em que suas premissas sejam verdadeiras e sua conclusão seja falsa. Caso essa combinação (de premissas verdadeiras e conclusão falsa) seja possível, teremos um contraexemplo que evidencia que o argumento não é válido.

Para ilustrar esse ponto, consideremos o seguinte argumento:

P₁: Se João está tocando bateria, então ele está no quarto.

P₂: João está no quarto.

C: Portanto, João está tocando bateria.

É possível arrumarmos um contraexemplo para o argumento em questão? Ou seja, é possível ter uma situação em que as premissas do argumento sejam verdadeiras e a sua conclusão seja falsa? De novo, não precisamos saber se de fato as premissas são verdadeiras. Podemos refletir com base em hipóteses ou suposições para analisar a validade do argumento, mesmo sem saber quem é João, se ele toca bateria no quarto etc. Suponhamos que **P₁** e **P₂** sejam verdadeiras. Disso se segue necessariamente que João está tocando bateria?

Considerando as premissas do argumento, é perfeitamente possível ter uma situação em que João esteja no quarto e, mesmo assim, não esteja tocando bateria. Com isso, temos um contraexemplo para o argumento. Portanto, de acordo com o critério de validade apresentado, o argumento não é válido.

Um argumento é **inválido** se e somente se não é válido; isto é, o argumento admite um contraexemplo no qual as premissas são verdadeiras e a conclusão falsa, ao mesmo tempo.

Tendo em vista as noções de validade e invalidade de argumentos, consideremos agora um argumento muito parecido com o argumento precedente, mas, com uma ligeira diferença, que faz com ele seja considerado válido.

P₁: Se João está tocando bateria, então, ele está no quarto.

P₂: João está tocando bateria.

C: Segue-se que João está no quarto.

O argumento é válido? Uma boa maneira para responder essa pergunta é refletir se há um contraexemplo para o argumento. Ou seja, seria possível que suas premissas fossem todas verdadeiras e, ao mesmo tempo, sua conclusão falsa? Consideremos o que acontece na situação em que as premissas são verdadeiras. Supondo que **P₁** seja verdadeira, então, João não está tocando bateria sem estar no quarto. Supondo, além disso, que **P₂** também seja verdadeira, então, João está tocando bateria. Agora, em tal situação, a conclusão, que diz que João está no quarto, poderia ser falsa? Como não pode haver situação em que, ao mesmo tempo, ambas as premissas sejam verdadeiras e a conclusão falsa, temos que o argumento é válido. Se as premissas forem verdadeiras, a conclusão necessariamente será verdadeira. Dizemos que argumentos válidos garantem logicamente sua conclusão. Mesmo sem saber nada sobre João, podemos perceber que as premissas nos obrigam, em um sentido lógico, a aceitar a conclusão.

O interessante para a consequência lógica não é o conteúdo informativo das premissas e da conclusão, e sim a forma ou estrutura do argumento. Para

tratar desse ponto, consideremos a forma do argumento que acabamos de examinar:

$$\begin{array}{c} \text{Se } \alpha, \text{ então, } \beta \\ \alpha \\ \therefore \beta \end{array}$$

As letras gregas aqui estão representando sentenças assertivas e o símbolo ‘ \therefore ’ serve para indicar a conclusão do argumento. Se α e β representarem, respectivamente, as sentenças ‘João está tocando bateria’ e ‘João está no quarto’, temos o último argumento analisado acima como uma instância do esquema em questão. Agora, se preservarmos essa mesma forma e substituirmos as letras α e β , respectivamente, pelas seguintes sentenças ‘Deus é perfeito’ e ‘Deus não está sujeito a carência alguma’, temos, como instância da mesma estrutura, o primeiro argumento que vimos na seção precedente.

Podemos providenciar muitos argumentos sobre vários assuntos, ao passo que, se a estrutura acima for preservada, a validade do argumento também será. Assim, podemos dizer que a validade dos argumentos, em um contexto lógico, não depende do conteúdo, e sim da forma ou estrutura do argumento.

Nesse sentido, não cabe à Lógica investigar a veracidade do conteúdo veiculado pelas afirmações contidas nos argumentos analisados. Não é uma questão lógica saber, por exemplo, se a sentença que diz que João toca bateria é de fato verdadeira. Uma investigação a respeito da verdade de tal sentença seguiria um procedimento empírico, e não um procedimento lógico. Analogamente, não é uma questão lógica investigar se a sentença que diz que Deus é perfeito é verdadeira ou não. Cabe a algum tipo de estudo teológico a investigação sobre tal afirmação. Podemos ter argumentos que tratam de diversos assuntos, e não cabe à Lógica investigar se suas premissas são de fato verdadeiras. Considerando que podemos ter argumentos sobre todo tipo de assunto (Biologia, Física, Matemática, Psicologia, etc.), caberia à Lógica investigar se as premissas dos diversos argumentos de que dispomos são verdadeiras? Se fosse tarefa da Lógica investigar, para cada afirmação que compõe um argumento, se é verdadeira ou não, a Lógica teria que tratar da totalidade do conhecimento humano, e isso definitivamente não é tarefa da Lógica. Além disso, é comum tratarmos de situações hipotéticas em nossos argumentos, nas quais não sabemos se as

premissas são verdadeiras, ou mesmo em contextos em que sabemos que as premissas são falsas e apenas queremos indicar o que se seguiria logicamente daquelas premissas, caso fossem verdadeiras.

De acordo com a noção de validade apresentada acima, se as premissas de um argumento forem verdadeiras, a conclusão necessariamente será verdadeira. Note que não é exigido que as premissas sejam de fato verdadeiras; apenas que, em um argumento válido, não pode haver uma situação em que as premissas sejam verdadeiras e conclusão falsa. Podemos, inclusive, ter argumentos válidos com uma ou mais afirmações falsas. Considere o seguinte exemplo:

P₁: Ou guitarras são animais de estimação, ou são instrumentos musicais.

P₂: Guitarras não são instrumentos musicais.

C: Logo, guitarras são animais de estimação.

Podemos dizer que o argumento é válido? Se as premissas forem verdadeiras (não estamos dizendo que de fato são), a conclusão teria necessariamente de ser verdadeira. Mesmo sabendo que **P₂** é falsa, podemos perguntar o que aconteceria com a conclusão, caso todas as premissas (incluindo **P₂**) fossem verdadeiras. Considerando esse caso em que as premissas seriam verdadeiras, a conclusão necessariamente seria verdadeira, embora a conclusão seja patentemente falsa. A conclusão é consequência lógica das premissas. Portanto, de acordo com a noção de validade vista acima, o argumento em questão é válido, mesmo tendo afirmações falsas. Note que esse argumento tem a mesma forma ou estrutura do primeiro argumento que apresentamos nesta seção (que tinha como conclusão a afirmação ‘João está na universidade’):

Ou α , ou β .

Não é o caso que α .

$\therefore \beta$

Como dissemos, não importa o conteúdo do argumento em questão, nem se as premissas são verdadeiras. Tendo isso em vista, podemos substituir

as variáveis (α e β) por sentenças que tratam de qualquer assunto, em situações reais, ficcionais, hipotéticas etc., ao passo que, se preservarmos a estrutura, preservaremos a validade do argumento.

Voltemos ao nosso exemplo em que temos um argumento válido, mesmo com afirmações falsas. Embora a conclusão seja consequência lógica das premissas, o argumento não servirá para sustentar a conclusão em uma discussão racional. Dizemos que embora o argumento seja válido, ele não é *correto* (ou *sólido*). Para que um argumento seja correto, além de ser válido, ele precisa *ter premissas verdadeiras*.

Um argumento é **correto** se e somente se:

- (i) ele é válido e
- (ii) suas premissas são verdadeiras.

Como um argumento, para ser correto, deve cumprir as condições (i) e (ii) do quadro acima, há três maneiras de termos argumentos que não são corretos: 1) o argumento pode ser válido sem ter todas as premissas verdadeiras; 2) o argumento pode ser inválido e ter premissas verdadeiras; 3) o argumento pode ser inválido e ter premissas falsas. Em qualquer desses casos, o argumento será incorreto.

Uma questão importante: é possível termos um argumento correto com conclusão falsa? Note que a definição de argumento correto acima fornecida não diz nada diretamente sobre a conclusão de um argumento correto. Mesmo assim, é fácil perceber que a conclusão de um argumento correto necessariamente deverá ser verdadeira. Suponhamos um dado argumento correto. Se ele é correto, ele será válido. Como o argumento é válido, se as premissas são verdadeiras, a conclusão será necessariamente verdadeira. Como o argumento em nossa suposição é correto, suas premissas são verdadeiras. Tendo isso em vista, a conclusão de qualquer argumento correto deverá ser verdadeira.

Outra questão que poderíamos colocar é: se um dado argumento tem premissas e conclusão verdadeiras, ele será correto? A resposta é não, pois ele pode ser inválido. Ou seja, um argumento pode ser inválido mesmo com premissas e conclusão verdadeiras. Tome o seguinte exemplo:

P1: Se São Luís é uma cidade brasileira, então, São Luís fica na América do Sul.

P2: São Luís fica na América do Sul.

C: São Luís é uma cidade brasileira.

As afirmações do argumento acima são todas verdadeiras; contudo, mesmo assim, ele não é correto. Particularmente, esse argumento é inválido. Argumentos inválidos, mesmo aqueles que tenham afirmações verdadeiras, não podem garantir a conclusão. Note que o argumento em questão tem a mesma forma ou estrutura do segundo argumento que tratamos nesta subseção (que tinha como conclusão a afirmação ‘João está tocando bateria’):

$$\begin{array}{c} \text{Se } \alpha, \text{ então, } \beta \\ \beta \\ \therefore \alpha \end{array}$$

Do mesmo modo que temos formas de argumentos que são válidas, temos formas de argumentos inválidas. Considerando esta estrutura, não importa quais sentenças colocamos no lugar das variáveis α e β , se preservarmos a forma inválida, o argumento resultará inválido e, portanto, incorreto. Mas, como saber se estamos diante de uma forma inválida de argumento? Bom, basta encontrarmos um contraexemplo. O contraexemplo pode ter qualquer conteúdo, desde que seja possível substituir as variáveis da forma indicada (preservando a sua estrutura) por uma situação (hipotética ou não) na qual as premissas sejam verdadeiras e a conclusão falsa.

Até aqui, vimos as noções de validade e correção de argumentos. Especificamente, apresentamos uma noção informal ou pré-teórica de consequência lógica como *preservação de verdade*: não é possível ter uma situação ou caso em que as premissas sejam verdadeiras e a conclusão seja falsa. Mas, o que é uma situação possível? Grande parte do empenho da Lógica foi tornar precisa essa noção de consequência lógica, inclusive tornando a ideia de “caso” ou “situação possível” mais clara. Com isso, no desenvolvimento da Lógica, surgiram várias lógicas ou teorias acerca da relação de consequência lógica de certos padrões de argumentos. Há várias lógicas que se diferenciam entre si

pelos seus princípios, métodos e campos de aplicação, e que, portanto, podem apresentar resultados distintos. Na seção 3, veremos como exemplo a lógica proposicional clássica e vamos conferir algumas formas ou padrões de argumentos que são considerados válidos em tal lógica. Antes disso, discutiremos, na próxima subseção, alguns tipos de argumentos.

2.3. Deduções, induções e abduções

Quando saímos do território puramente formal ou esquemático da Lógica e nos preocupamos com suas aplicações no contexto dos argumentos cotidianos, entramos no território usualmente chamado de *lógica informal*. Aqui, estamos interessados em outros tipos de consequências argumentativas (além daquela *consequência lógica* definida antes como paradigmática), os quais também são úteis para a Filosofia e as ciências. A lógica informal também se preocupa as diferentes categorias de *pseudoargumentos* ou argumentações falaciosas.

Uma primeira distinção importante: nem todo argumento correto (do tipo descrito na seção precedente) é um bom argumento para ser usado em um debate racional, assim como nem todo argumento bom em um debate racional precisa se comportar como um argumento correto!

Como seria, então, um argumento bom ou proveitoso para ser usado em um debate racional? Antes, precisamos distinguir as categorias mais gerais de argumentos, para, então, indicar algumas exigências mínimas que os argumentos de cada categoria precisam satisfazer para serem considerados bons. Essas categorias mais gerais de argumentos são três, a saber: os dedutivos, os indutivos e os abductivos.

Já conhecemos o tipo mais importante de todos, considerado a *inferência lógica por excelência*, a mais segura e perfeita de todas: os argumentos **dedutivos**. São exatamente os mesmos que foram classificados como **válidos** de um ponto de vista estritamente lógico. Há uma sutil diferença conceitual entre as duas nomenclaturas; porém, como elas se equivalem (todo argumento logicamente válido é dedutivo, e vice-versa), não nos estenderemos nesse ponto aqui.

Como você pode distinguir de maneira segura entre deduções e outros tipos de inferências? Ao se deparar com um argumento qualquer, primeiro identifique sua estrutura (premissas e conclusão), aplicando as indicações fornecidas na seção precedente. Daí, você deve se perguntar o seguinte:

Supondo que todas as premissas do argumento sejam verdadeiras, seria impossível que a conclusão fosse falsa? Caso afirmativo, o argumento é uma dedução; caso negativo, não é dedução.

Quando mencionamos “impossível” aqui, não confunda com “improvável”. Existem situações altamente improváveis, mas que não são impossíveis *de um ponto de vista lógico*. Alguém que vivesse na Idade Média poderia achar impossível que pudéssemos enviar sondas para observar de perto outros planetas, mas isso era apenas uma crença falsa. O que estamos chamando de impossível aqui é algo bem mais forte, e tem a ver com o que é literalmente *absurdo* ou *inconcebível* – por exemplo, que um triângulo tenha quatro vértices.

Por outro lado, um argumento **indutivo** é aquele *cujas hipóteses nos induzem a aceitar sua conclusão, mas apenas com um caráter probabilístico*. É como se uma indução nos apresentasse fortes indícios para aceitar sua conclusão, mas ainda assim, sempre existe a *possibilidade lógica* de que a conclusão seja falsa. Não há garantia absoluta de que a conclusão deva ser aceita, uma vez consideradas as premissas. Por isso, embora as induções desempenhem um papel muito importante na construção do conhecimento, do ponto de vista estritamente lógico, as induções *sempre* podem produzir conclusões falsas (mesmo que, em alguns casos, isso seja altamente improvável ou até fisicamente impossível).

Uma **indução** é uma inferência cujas premissas sugerem fortemente sua conclusão, sem, contudo, garanti-la logicamente.

Seguem alguns exemplos de argumentos indutivos, cuidadosamente escolhidos para ilustrar algum ponto relevante:

a)

P₁: *Fumantes compulsivos tendem a desenvolver algum tipo de câncer.*

P₂: *Fulano é fumante compulsivo.*

C: *Logo, fulano provavelmente desenvolverá algum tipo de câncer.*

b)

P₁: *Todos os indivíduos daquela espécie observados até agora pelos cientistas apresentaram tal comportamento.*

C: *Portanto, todos os indivíduos daquela espécie apresentam tal comportamento.*

c)

P₁: *As estatísticas mostraram uma relação diretamente proporcional entre o aumento do desemprego e o aumento da violência nos centros urbanos.*

C: *Nos casos observados, o desemprego pode, pois, ser considerado uma causa para a violência urbana.*

Um detalhe importante: a definição antiga, e ainda muito citada, para argumentos indutivos consiste em dizer que eram inferências que “de premissas particulares, derivavam conclusões gerais”. Nossos exemplos mostram que nem sempre é assim! Apenas o item (b) apresenta uma generalização (no caso, para todos os indivíduos) a partir de observações parciais (os indivíduos observados até o momento), e ainda assim, a própria premissa já é geral (porque fala de *todos* os indivíduos observados). O item (c) pretende encontrar uma relação causal entre dois eventos gerais (o aumento do desemprego e o aumento da violência urbana), e o item (a) chega a fazer justo o inverso do que pretende aquela noção tradicional, pois de uma premissa geral deriva uma conclusão particular!

Tradicionalmente, os lógicos apenas falavam de deduções e induções, mas uma modalidade de raciocínio, bastante empregada na investigação científica, não se encaixava exatamente em nenhuma das duas modalidades. Daí, pesquisadores como Charles Sanders Peirce aprimoraram a compreensão acerca dessa forma de raciocinar, denominando-a de abduções. A *abdução* (ou *retrodução*) ficou conhecida também como a *inferência da melhor explicação disponível*:

Uma **abdução** (ou **retrodução**) é uma inferência que remonta dos efeitos às suas (prováveis) causas, selecionando a melhor explicação disponível

Em outras palavras, quando você precisa explicar um fato B, já bem-sabido e estabelecido, e dispõe de uma hipótese explicativa A (ou um conjunto consistente de hipóteses) que explica suficientemente bem o fato B, e, além disso, se A for verdadeira, seria a melhor explicação disponível para B, você pode considerar A como verdadeira (pelo menos até que se disponha de uma explicação ainda melhor, ou até que novas informações acerca de B nos obriguem a encontrar outra explicação).

É importante mencionar que a abdução, como toda inferência, é um procedimento para obtenção de novas informações a partir das informações disponíveis. Por isso, abduções dependem de verificações independentes, porque, assim como as induções, também podem produzir conclusões equivocadas.

Por exemplo, suponha que você seja um estudante e divida o apartamento com um colega, apenas os dois têm as chaves do lugar e nenhum dos dois costuma receber visitas sem avisar ao outro. Suponha, além disso, que você tenha passado o dia fora, chegou tarde da noite e encontrou a metade de uma pizza guardada no forno (e não foi você quem comprou a dita cuja). Qual sua conclusão mais natural? Foi seu colega quem a comprou. Isso foi uma abdução. Poderia ter acontecido outra coisa? Sim, mas dadas as circunstâncias e hábitos dos moradores, essa era a explicação mais plausível para o acontecimento. Para ter certeza, você poderia fazer uma verificação independente acerca de sua conclusão e procurar outros indícios, ou – melhor ainda – perguntar ao seu colega! Mas essa verificação independente precisou que você primeiro fizesse a abdução, para ter um direcionamento investigativo (uma direção para a qual olhar em busca de respostas).

As abduções desempenham um papel muito importante na investigação científica. O cientista se defronta com uma determinada classe de fenômenos e, apoiado no conhecimento que possui sobre seu objeto de estudos, emprega sua criatividade para elaborar modelos explicativos para aqueles fenômenos. Daí, ele escolhe o modelo explicativo que melhor se aplica e busca evidências novas que o corroborem. Exemplos muito interessantes de abduções podem ser encontrados em ciências como a biologia evolucionista e a cosmologia (parte da Física que estuda a estrutura e origem do universo), mas não são fáceis de serem descritos em poucas palavras, porque envolvem modelos bastante complexos.

Porém, dê uma olhada nesses exemplos simples de abduções, para ilustrar melhor o assunto:

a)

P₁: *Nesta floresta é muito comum encontrar onças.*

P₂: *Essas pegadas são compatíveis com felinos de grande porte.*

C: *Logo, essas pegadas (provavelmente) são de onça.*

b)

P₁: *Em um depósito de alimentos, cada espécie de cereal está armazenada em um saco diferente.*

P₂: *Encontrei alguns feijões espalhados pelo chão.*

C: *Esses feijões no chão (provavelmente) estavam no saco de feijões.*

c)

P₁: *Está havendo uma epidemia de conjuntivite na escola de meu filho.*

P₂: *Meu filho apresentou sintomas de conjuntivite após voltar da aula.*

C: *Logo, ele (provavelmente) contraiu conjuntivite em sua escola.*

Se você costuma assistir a filmes ou séries policiais envolvendo a investigação de crimes, irá se dar conta de que os investigadores estão quase todo o tempo fazendo abduções, além de algumas poucas deduções e induções para arrematar suas teorias. Portanto, embora o famoso personagem fictício Sherlock Holmes (criado por Arthur Conan Doyle) costumasse descrever seu método para o Dr. Watson como sendo dedutivo, a rigor ele estava principalmente realizando abduções, não deduções!

As abduções se parecem muito com as induções, não é mesmo? Tanto que só recentemente foram classificadas como categorias distintas de argumentos. Talvez fique mais claro se reservarmos o rótulo de “indutivos” para aqueles argumentos nãodedutivos que estão tentando nos convencer de que algo acontece (ou irá acontecer), enquanto os argumentos abduativos tentam nos convencer de quais foram as causas de algo que já sabemos ser o caso – daí, o apelido de “inferência da melhor explicação”.

Outro ponto interessante: pela sua forma, os argumentos abduativos se assemelham a um tipo bastante comum de falácia (ver próxima seção), denominada como “afirmação do conseqüente”, cuja estrutura já foi usada na seção precedente como exemplo de estrutura inválida:

$$\begin{array}{l} \text{Se } \alpha, \text{ então, } \beta \\ \beta \\ \therefore \alpha \end{array}$$

Porém, essa semelhança pode ser contestada com uma análise mais aprofundada. De qualquer maneira (e para evitar complicações desnecessárias nesta exposição introdutória), é preciso enfatizar o caráter *heurístico* (investigativo) das abduções – como em nosso exemplo da pizza encontrada no forno. Afinal, elas não provam efetivamente que algo aconteceu – no caso, que o colega de quarto comprou a pizza –, mas direcionam a pesquisa em determinada direção com o objetivo de coletar mais e mais evidências, fazer predições testáveis a partir da conclusão obtida etc. (Mas, esse assunto nos leva para fora da Lógica e na direção da Filosofia da Ciência.)

Finalmente, para encerrar esta seção, indicaremos, em linhas gerais, quando um argumento deve ser considerado *bom para uma discussão racional*. Isso depende de se estamos diante de argumentos válidos (dedutivos) ou inválidos (indutivos e abduativos).

Se o argumento em questão for válido, para ser um bom argumento, ele precisa ser correto (na acepção explicada anteriormente) e suas *premissas precisam ser mais garantidas, por assim dizer, do que sua conclusão*. Em outras palavras, a conclusão não pode ser redundante em relação às premissas.

Consideremos o seguinte exemplo de um argumento correto (válido com premissas verdadeiras), porém, inútil para uma investigação ou discussão visando aprimorar nosso conhecimento:

P₁: Não acontece que ácidos não tenham o PH menor do que 7.

C: Logo, ácidos têm o PH menor do que 7.

Por sua vez, quando devemos considerar argumentos indutivos e abdu-
tivos como bons argumentos (já supondo que suas premissas sejam verdadeiras)?
É claro que argumentos inválidos, em geral, não garantem necessariamente sua
conclusão; porém alguns são considerados melhores do que outros. No caso dos
indutivos, isso depende da *força* com que a conclusão é obtida. Em outras pala-
vras: *quanto mais provável for a conclusão do argumento indutivo, tanto mais
forte ele é*. O problema maior é determinar essa probabilidade, e isso depende
de cada situação (o que é provável em um contexto ou a partir de certo conjunto
de dados pode ser altamente improvável em outro contexto).

E no caso das abduções? O que ocorre é que, como explicamos, estas
funcionam como um procedimento heurístico (investigativo) para escolher a
melhor explicação para um fenômeno. Contudo as conclusões obtidas abdu-
tivamente precisam ser confrontadas com verificações independentes; assim, sua
importância consiste principalmente em fornecer boas direções investigativas, ao
longo das quais você coletará mais dados experimentais, fará novas previsões
indutivas, etc.

2.4. Algumas falácias

Antes de encerrarmos nosso breve passeio pela lógica informal, con-
versaremos rapidamente sobre um problema muito grave em todo e qualquer
debate intelectualmente honesto (não apenas os de natureza acadêmica): o uso
de argumentos falaciosos.

As falácias (ou **sofismas**) consistem em argumentos ilegítimos, nos quais
as premissas não acarretam (nem mesmo indutivamente) a suposta con-
clusão, embora pareçam fazê-lo.

Com a definição acima, tentamos capturar duas características impor-
tantes de qualquer falácia: seu caráter ilegítimo e ao mesmo tempo sedutor. As
falácias são, portanto, erros argumentativos, os quais são apresentados como se
fossem inferências legítimas (boas), e que com frequência conseguem enganar
alguns interlocutores desatentos.

O erro fundamental de todas as falácias é sempre o mesmo: suas premissas são *irrelevantes* para a conclusão, mesmo quando pareçam ser relevantes. Em outras palavras, a conclusão simplesmente *não é consequenciado* que é afirmado nas premissas. Porém nem sempre é fácil diagnosticar essa falha em uma leitura superficial ou quando estamos distraídos por algum motivo (por exemplo, estando sob intensa emoção ou se temos algum forte interesse pessoal no assunto debatido).

Nem sempre uma falácia é empregada com má-fé. A má-fé acontece somente quando aquele que enuncia o argumento falacioso sabe que está cometendo um erro argumentativo, mas espera mesmo assim enganar sua plateia. Também pode acontecer de uma falácia ser usada sem que se tenha consciência disso, apenas por descuido, limitação cognitiva ou simples ignorância. Porém, uma vez que o erro argumentativo foi percebido ou denunciado, é inaceitável que ele continue sendo sequer considerado.

Qualquer indivíduo com habilidades cognitivas e lógicas medianas consegue perceber o problema com a maioria das falácias, desde que elas sejam examinadas com a devida atenção. Conhecer e estudar antecipadamente uma lista razoável de tipos de falácias é, portanto, muito útil para qualquer pessoa, visto que todos nós precisamos em algum momento defender ou criticar racionalmente uma posição, seja nossa, seja de terceiros.

Outro ponto importante que você deve entender é que, ao diagnosticar uma falácia, não importa se a conclusão obtida é verdadeira ou falsa, porque é *a maneira como ela foi obtida* que está errada. Aquela conclusão pode até ser verdadeira (independentemente da falácia usada para defendê-la), mas não tem como ser sustentada daquela maneira, cabendo ao seu proponente construir outro argumento, desta vez legítimo, para sustentar sua posição.

Por esse motivo, desde que Aristóteles sistematizou pela primeira vez o estudo da lógica, há cerca de 2.400 anos, até os nossos dias, as falácias têm sido estudadas e catalogadas para garantir uma boa argumentação na Filosofia e nas ciências. Assim, dezenas e dezenas de tipos e estruturas falaciosas diferentes vêm sendo listadas nos manuais especializados, com nomenclaturas tradicionais e modernas, para facilitar uma rápida identificação em um contexto prático.

Devido à limitação de espaço, apresentaremos apenas alguns tipos famosos de falácias, selecionando-as entre as que são empregadas com bastante

frequência em conexão com a pesquisa científica. Para um estudo mais abrangente, recomendamos as leituras listadas no final deste capítulo.

Algumas falácias podem ser reconhecidas tão somente observando a estrutura do pseudoargumento. Observe os dois seguintes exemplos:

a)

P₁: *Caso a data em questão caia num sábado, não haverá aula.*

P₂: *Não haverá aula naquela data.*

C: *Logo, a data cairá em um sábado.*

b)

P₁: *A música popular tem características inferiores à música erudita.*

C: *Logo, a música erudita é claramente superior à música popular.*

Na letra (a), temos uma estrutura argumentativa do tipo:

A implica em B. B acontece. Logo, A acontece.

Essa estrutura é claramente injustificada, porque B pode acontecer por outro motivo – no caso de nosso exemplo, pode ser que seja um feriado! Não há nenhuma razão para concluir que A aconteceu ou acontecerá (pois, tanto pode ser o caso como não ser o caso). Esse tipo de falácia é conhecido como *afirmação do consequente*, porque uma das premissas é um enunciado condicional (uma sentença da forma: “se A, então, B”, onde A é o antecedente do condicional, e B é o consequente do condicional).

Caso a data em questão caia num sábado, não haverá aula.

A → **B**

Não haverá aula naquela data.

B

Logo, a data cairá em um sábado.

A

A estrutura falaciosa em questão pode, pois, ser definida assim:

a) A **falácia de afirmação do consequente** ocorre quando, a partir de um enunciado condicional tomado como premissa, e da afirmação de seu consequente, pretende-se inferir o antecedente do condicional.

Nós já discutimos, na seção 2.2, o problema com essa estrutura, quando examinamos o seguinte exemplo de argumento:

P₁: Se João está tocando bateria, então ele está no quarto.

P₂: João está no quarto.

C: Portanto, João está tocando bateria.

Na letra (b), temos outro caso de falácia que pode ser reconhecida tão somente examinando sua estrutura, que é ainda mais simples:

A; portanto B. B; portanto C. C; portanto, D. (...); portanto, A.

Trata-se do famigerado círculo vicioso, assim denominado porque a conclusão apenas repete uma das premissas, fazendo uma espécie de círculo. Em nosso exemplo, temos o caso mais extremo: apenas uma premissa e uma conclusão (a estrutura resulta assim: 'A, portanto A'). Não se deixe enganar pelo

palavrado ligeiramente diferente, pois a conclusão está apenas repetindo a premissa. Já explicamos anteriormente que argumentos redundantes não são *bons* argumentos, ainda que sejam válidos!

A música popular tem características inferiores à música erudita.

A

Logo, a música erudita é claramente superior à música popular.

A

A estrutura falaciosa em questão pode, pois, ser definida assim:

b) A **falácia do círculo vicioso** ocorre sempre que a conclusão do argumento simplesmente repete uma das premissas.

Os próximos exemplos ilustrarão outros tipos muito comuns de argumentações falaciosas.

c)

Deve existir alguma verdade na astrologia, porque milhões de pessoas em todo o mundo acreditam em horóscopos e pautam suas decisões assim!

d)

Nunca se conseguiu mostrar que existe vida fora da Terra. Portanto, não existe tal coisa.

e)

Não levo a sério essas regulações do Ministério da Saúde contra o tabagismo. Certamente são falsas. Meu avô fumou a vida inteira e morreu de causas naturais com quase cem anos de idade!

Os pseudoargumentos acima são, respectivamente, casos típicos das seguintes modalidades de falácias (tente perceber a correspondência):

c) A **falácia do recurso à quantidade** ocorre sempre que se pretende defender (ou rejeitar) alguma posição tão somente valendo-se do grande número de indivíduos que defendem (ou rejeitam) aquela posição.

A falha por trás desse raciocínio deveria ser óbvia: não é quantidade de aderentes a uma tese que faz com que essa tese seja verdadeira, nem a quantidade de pessoas que a rejeitam que faz com que ela seja falsa. Fatos científicos não dependem de que se acredite neles para serem cientificamente bem estabelecidos. Em certo momento (por volta do século XV), praticamente toda a humanidade acreditava que o Sol girava em torno da Terra, e todos estavam errados, como hoje sabemos!

d) A **falácia do recurso à ignorância** ocorre sempre que se pretende usar tão somente a ausência de provas em favor de (ou contra) uma tese como sendo prova suficiente de que essa tese é verdadeira (ou falsa).

Obviamente, a ausência de provas não pode ser considerada como uma prova em si mesma. Só porque algo não foi provado (ainda), não quer dizer que seja falso. É claro que ninguém é obrigado a aceitar algo sem evidências, mas também não pode logicamente estar certo de que essa informação é falsa só porque ainda não foram produzidas provas em seu favor. A posição mais razoável é a suspensão do juízo, até que se colem evidências seguras. (Uma exceção honrosa pode ser encontrada na atividade jurídica, onde, por razões óbvias, um réu é inocente até que se prove sua culpa – a ausência de provas de culpabilidade deve obrigatoriamente ter o mesmo efeito prático de uma prova de sua inocência.)

e) A **falácia da evidência anedótica** ocorre sempre que se pretende defender uma posição baseando-se em relatos pessoais isolados, sem atender a critérios objetivos de verificação.

Essa falácia mereceria uma análise detalhada se dispuséssemos de mais tempo, porque aparentemente muitas pessoas não conseguem perceber seu problema. Porém, deve ser suficiente mencionar que existem diversas boas razões para que ela seja rejeitada em uma discussão acadêmica, razões que passam pela não confiabilidade de certos relatos pessoais – há estudos científicos muito interessantes sobre falsas memórias –, sem falar nas mentiras inocentes e nos exageros, na mitomania (a condição psicológica quando o mentiroso acredita piamente na própria mentira), e o conflito entre uma evidência científica e uma evidência informal (sem nenhum rigor na coleta de dados, controle das variáveis envolvidas etc.). Além disso, relatos isolados, mesmo quando verdadeiros, não têm força suficiente para desmentir estatísticas cuidadosamente obtidas. Em nosso exemplo, o fato de aparentemente alguns tabagistas não desenvolverem doenças ligadas ao seu vício não torna falsas as conclusões dos estudos *estatísticos* correlacionando os dois fatores (tabagismo e certos grupos de doenças) em larga escala.

3. O Cálculo Proposicional Clássico (CPC)

Nesta seção, apresentaremos uma teoria lógica muito simples conhecida como *Cálculo Proposicional Clássico* (CPC). A partir do CPC, veremos como é possível verificar, de maneira precisa, a validade de uma gama de argumentos que podem ser expressos nessa teoria. É interessante notar que não pretendemos oferecer uma apresentação detalhada do CPC. Nosso interesse com esta seção, além de termos uma amostra de uma teoria lógica bastante simples, consiste em conferir um tratamento formal para a noção de consequência lógica que, até então neste texto, foi tratada de maneira informal ou pré-teórica. Para tanto, na subseção 3.1, vamos apresentar uma linguagem para o CPC que será usada para a apresentação do CPC em uma abordagem semântica na subseção 3.2, e para uma abordagem sintática na subseção 3.3.

3.1. A Linguagem do CPC

Como mencionamos na introdução deste texto, com o seu desenvolvimento, a Lógica passa a ser qualificada como simbólica (ou matemática), sobretudo, por fazer uso de linguagens formais (ou artificiais). A vantagem de termos uma linguagem formal é que poderemos *traduzir*, por assim dizer, sentenças originalmente em uma dada linguagem natural, como o português, para uma linguagem artificial, cuja estrutura está precisamente especificada. Como, a rigor, esse processo não é uma tradução, ele é chamado frequentemente de *formalização* (ou *simbolização*). Com linguagens formais, podemos evitar alguns problemas de imprecisão, como ambiguidade e vagueza, que são próprios das línguas naturais. Embora ambiguidade e vagueza contribuam para a flexibilidade das línguas naturais, elas são indesejáveis para um tratamento formal da noção de consequência lógica.

Nesta subseção, veremos uma linguagem para o CPC ou simplesmente uma linguagem proposicional. Para definir uma linguagem proposicional, iremos adotar o seguinte procedimento: i) apresentar o alfabeto, caracterizando os símbolos que vão compor a linguagem; ii) apresentar um conjunto de regras, caracterizando aquilo que são fórmulas (expressões gramaticamente bem formadas) da linguagem. Definiremos uma linguagem proposicional, denominada aqui de **L**, contendo em seu alfabeto *variáveis proposicionais*, *conectivos lógicos* e *sinais de pontuação*.

As variáveis proposicionais serão letras maiúsculas do alfabeto latino, usadas para representar (ou simbolizar) sentenças simples como ‘João está na universidade’, ‘São Luís é uma cidade brasileira’, ‘ácidos tem o PH maior do que 7’. Para não restringir **L** apenas às 26 letras do alfabeto latino, iremos fazer uso de subscritos numéricos para garantir um estoque infinito de variáveis proposicionais: $A, B, \dots, Z, A_1, B_1, \dots, Z_1, A_2, \dots, Z_2, \dots$. Como o conjunto dos números naturais é infinito, temos assim um conjunto infinito enumerável de variáveis proposicionais. As variáveis proposicionais poderão ser conectadas para representar sentenças complexas como ‘João é estudante de filosofia e músico’, ‘Pedro irá para universidade ou para a praia’, ‘se João está no quarto, então está tocando bateria’ etc.

Nos exemplos acima, sentenças simples estão sendo conectadas pelos conectivos ‘e’, ‘ou’ e ‘se..., então’ que temos em português. Os correspondentes

aos conectivos do português, em uma linguagem proposicional, são os operadores (ou conectivos) lógicos. Em **L** faremos o uso dos seguintes operadores lógicos: ' \neg ', ' \wedge ', ' \vee ', ' \rightarrow ', ' \leftrightarrow '. Vamos conferir no quadro abaixo o nome que usualmente é dado para cada um desses operadores lógicos, bem como algumas expressões do português que eles representam.

Operador Lógico	Nome do Operador	Conectivos do Português
\neg	Negação	'não', 'não é o caso que', 'é falso que'.
\wedge	Conjunção	'e', 'mas', 'porém'.
\vee	Disjunção	'ou', 'e/ou', 'ou ..., ou'.
\rightarrow	Implicação	'se... então...'
\leftrightarrow	Bi-implicação	'se e somente se', 'é equivalente a'.

O operador de implicação material, ' \rightarrow ', é chamado também de condicional material e, analogamente, o operador de bi-implicação material, ' \leftrightarrow ', é chamado também de bicondicional material. Cabe ressaltar que a lista de conectivos do português no quadro acima não é exaustiva. Ela traz apenas alguns exemplos, para termos um significado intuitivo de cada um dos operadores que estão sendo apresentados aqui. O significado formal de cada um dos operadores será dado na próxima subseção deste capítulo.

Além das variáveis proposicionais e dos operadores lógicos, a linguagem **L** também é dotada de sinais de pontuação, que têm o papel de evitar ambiguidades na linguagem. Os sinais de pontuação de **L** são os símbolos de parêntese abrindo, '(' , e de parêntese fechando, ') '.

DEFINIÇÃO: O **alfabeto** de **L** consiste nos seguintes símbolos:

i) um conjunto enumerável de variáveis proposicionais:

$A, B, \dots, Z, A_1, B_1, \dots, Z_1, A_2, \dots, Z_2, \dots$

ii) um conjunto de conectivos (ou operadores) lógicos: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$.

iii) Sinais de pontuação: (,).

Com base nesse alfabeto, uma *expressão* de **L** é entendida como qualquer sequência finita de elementos do alfabeto de **L**. Assim, em **L** temos

expressões como, por exemplo, ' $F \wedge M$ ', ' $U \vee P$ ', ' $A \vee \vee BU \leftrightarrow$ ' ou ' $\neg PP \wedge$ '. Nesses exemplos, apenas as duas primeiras expressões, como veremos, contam como expressões gramaticalmente bem formadas de L . As duas últimas do exemplo violam regras de formação de L . As expressões bem formadas são chamadas simplesmente de *fórmulas*. Vejamos, então, a definição de fórmula a partir das regras de formação da linguagem L .

DEFINIÇÃO: Uma **fórmula** de L é uma expressão que pode ser obtida através das seguintes regras:

- i) uma variável proposicional sozinha é uma fórmula;
- ii) se α é um fórmula, então $\neg\alpha$ é uma fórmula;
- iii) Se α e β são fórmulas, então $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, $(\alpha \rightarrow \beta)$ e $(\alpha \leftrightarrow \beta)$ são fórmulas;
- iv) Nada mais é uma fórmula.

Note que as expressões que estão na definição de fórmula acima, como ' $\neg\alpha$ ' e ' $(\alpha \wedge \beta)$ ', não são fórmulas de L . A rigor, elas não são expressões de L , já que as letras do alfabeto grego, α , β , etc., não constam no alfabeto de L . (Lembre-se de que as variáveis proposicionais de L são letras maiúsculas do alfabeto latino, e não do alfabeto grego). Como toda fórmula L é, por definição, uma expressão de L , seqüências de símbolos que não são expressões de L automaticamente não serão fórmulas de L . Na definição acima, as letras do alfabeto grego estão sendo usados como *metavariáveis*; elas não estão na linguagem que estamos tratando (linguagem-objeto L), e sim em uma linguagem que estamos usando para tratar de L (a *metalinguagem* de L).

As fórmulas de L sem operadores lógicos, conforme a cláusula (i) da definição de fórmula, são chamadas de *fórmulas atômicas*; elas correspondem a sentenças simples das línguas naturais. As fórmulas de L que são construídas a partir de sentenças atômicas conectadas através dos operadores lógicos, conforme as cláusulas (ii) e (iii) da definição de fórmula, são chamadas de *fórmulas moleculares*; elas correspondem a sentenças complexas das línguas naturais. A cláusula (iv) da definição de fórmula, conhecida como cláusula de fechamento, é importante para garantir que nenhuma expressão, com uma forma distinta das

cláusulas precedentes, seja construída. Além de impedir que expressões de L como ' $A \vee \vee BU \leftrightarrow$ ' e ' $\neg PP \wedge$ ' sejam consideradas fórmulas, a cláusula de fechamento é um componente importante para o tipo de definição que foi dada.

A definição de fórmula acima é uma definição *recursiva*. O conjunto de fórmulas de L foi definido recursivamente: primeiro, foram identificados os elementos iniciais (as fórmulas atômicas de L dadas pela cláusula (i) da definição); depois, foram listadas as condições que permitem que certas expressões sejam consideradas como pertencentes ao conjunto, se elas satisfazem as condições para tanto; por fim, a cláusula de fechamento impede que alguma expressão possa pertencer ao conjunto sem que sua inclusão seja exigida pelos passos anteriores da definição. A partir de tal definição recursiva, é sempre possível determinar, de maneira exata, para qualquer expressão de L , independentemente da complexidade da expressão, se ela pertence ou não ao conjunto das fórmulas de L .

DEFINIÇÃO: A **linguagem proposicional L** consiste no conjunto de todas as fórmulas de L .

3.2. Uma semântica para o CPC e a noção semântica de consequência lógica

Uma vez definida a linguagem para o CPC, nos ocuparemos com o fornecimento de uma semântica para essa linguagem. De maneira breve, chamamos de semântica *o estudo da interpretação das expressões de uma linguagem*. Estamos usando aqui o termo 'interpretação' em um sentido mais pontual ou técnico (em vez do sentido mais amplo adotado, por exemplo, na semântica filosófica), com a aceção de regras para associar cada unidade significativa da linguagem-objeto (aquela que está sendo estudada) com alguma referência (um objeto, uma propriedade, um conjunto de objetos, etc.).

A semântica padrão usada para o CPC é extremamente simples (por considerar apenas duas referências para as fórmulas: '1' ou '0' – intuitivamente: o *verdadeiro* ou o *falso*), e nos diz como calcular a referência (ou 'valor') de cada fórmula a partir de uma valoração *para a linguagem*. Intuitivamente, cada valoração para a linguagem faz algo bem preciso: atribui a cada fórmula atômica (ou

seja, cada variável proposicional) da linguagem exatamente um valor no conjunto $\{1,0\}$. A partir daí, a semântica clássica determina como calcular o valor de qualquer fórmula que contenha aquelas fórmulas atômicas. Vejamos rapidamente como isso é feito.

Primeiro, definimos o que é uma valoração:

DEFINIÇÃO: Uma **valoração** v é uma função que associa a cada fórmula atômica de **L** exatamente um elemento do conjunto $\{1, 0\}$ de valores de verdade.

Agora, definimos como obter o valor para qualquer fórmula de **L** a partir de uma valoração para as fórmulas atômicas de **L**. Para simplificar, usaremos a mesma notação v , embora, a rigor, estejamos definindo uma nova função interpretação (que determina o valor de *todas* as fórmulas com base em uma valoração específica para as fórmulas atômicas).

DEFINIÇÃO: Seja uma **valoração** v para as fórmulas atômicas de **L**. Assim, podemos estender v para determinar o valor de qualquer fórmula de **L** da seguinte maneira:

- i) $v(\neg\alpha) = 1$ se e somente se $v(\alpha) = 0$;
- ii) $v(\alpha \wedge \beta) = 1$ se e somente se $v(\alpha) = 1$ e $v(\beta) = 1$;
- iii) $v(\alpha \vee \beta) = 1$ se e somente se $v(\alpha) = 1$ ou $v(\beta) = 1$;
- iv) $v(\alpha \rightarrow \beta) = 1$ se e somente se $v(\alpha) = 0$ ou $v(\beta) = 1$;
- v) $v(\alpha \leftrightarrow \beta) = 1$ se e somente se $v(\alpha) = v(\beta)$.

Para ilustrar com um exemplo de que maneira as noções acima são aplicadas, consideremos novamente a linguagem **L** e uma valoração específica v_0 bem peculiar que definiremos assim:

$$v_0(A) = 1, v_0(B) = 1 \dots, v_0(Z) = 1, v_0(A_1) = 1, v_0(B_1) = 1, \dots, v_0(Z_1) = 1, \dots$$

(Ou seja, todas as fórmulas atômicas são verdadeiras na valoração v_0 .) Com base em v_0 , conseguimos saber o valor de qualquer fórmula de \mathbf{L} . Como exemplos de aplicação das regras acima, vejamos como calcular o valor das seguintes duas fórmulas *i)* $\neg\neg\neg A$ e *ii)* $\neg(A \rightarrow \neg B)$:

$$i) v_0(A) = 1 \Rightarrow v_0(\neg A) = 0 \Rightarrow v_0(\neg\neg A) = 1 \Rightarrow v_0(\neg\neg\neg A) = 0;$$

$$ii) v_0(A) = 1, v_0(B) = 1 \Rightarrow v_0(\neg B) = 0 \Rightarrow v_0(A \rightarrow \neg B) = 0 \Rightarrow v_0(\neg(A \rightarrow \neg B)) = 1.$$

Agora, chegamos numa parte realmente interessante. A semântica definida acima permite que calculemos também todos os valores de verdade possíveis para quaisquer fórmulas de \mathbf{L} , considerando *todas as valorações que poderiam ser fornecidas!* Esse procedimento, bastante conhecido, é denominado de *tabelas de verdade*.

Suponha que desejemos saber o valor daquela fórmula $\neg\neg\neg A$ usada há pouco; porém, não sabemos qual valoração específica considerar. Então, consideraremos todas as valorações v possíveis. Obviamente, há infinitas valorações para \mathbf{L} ; contudo, como a única fórmula atômica usada para construir a fórmula $\neg\neg\neg A$ foi a variável proposicional A , não faz sentido nos preocuparmos com aquelas infinitas valorações. Como existem apenas dois valores possíveis (1 e 0), ou cada valoração específica v' nos dará $v'(A) = 1$, ou nos dará $v'(A) = 0$. Construímos, então, uma tabela para determinar o que acontece com $\neg\neg\neg A$ em cada um desses dois casos possíveis:

A	$\neg A$	$\neg\neg A$	$\neg\neg\neg A$
1	0	1	0
0	1	0	1

Na primeira linha da tabela, colocamos o passo a passo da composição da fórmula desejada, a partir das fórmulas atômicas que nela aparecem; e, em cada linha abaixo, colocamos os valores calculados a partir de uma valoração possível para as fórmulas atômicas. Cada linha da tabela corresponde, portanto, a uma valoração possível para aquele conjunto de fórmulas atômicas e seus

consequentes valores para as fórmulas moleculares que as contêm, até descobrirmos o valor da fórmula desejada.

O procedimento é sempre automático, embora possa ficar cada vez mais demorado. Por exemplo, para construirmos a tabela de verdade para $\neg(A \rightarrow \neg B)$, precisaremos considerar quatro valorações possíveis para suas duas fórmulas atômicas.

A	B	$\neg B$	$A \rightarrow \neg B$	$\neg(A \rightarrow \neg B)$
1	1	0	0	1
0	1	0	1	0
1	0	1	1	0
0	0	1	1	0

A quantidade de combinações possíveis vai aumentando exponencialmente com a quantidade de fórmulas atômicas diferentes contidas na fórmula desejada. A quantidade de combinações é sempre 2^n , onde n é a quantidade de fórmulas atômicas diferentes. Por exemplo, na primeira tabela acima, só 1 fórmula atômica nos interessava, então precisamos de 2 linhas ($2^1 = 2$); na segunda tabela, eram 2 fórmulas atômicas iniciais (A e B), então, tivemos que usar 4 linhas ($2^2 = 4$); e assim por diante.

O emprego de tabelas de verdade (ou algum procedimento equivalente) nos permite identificar três categorias semânticas de fórmulas de **L**: as *contingentes* (fórmulas que podem receber o valor 1 ou 0 dependendo da valoração considerada), as *contraditórias* (que sempre recebem o valor 0 em todas as valorações possíveis) e as *tautológicas* (que sempre recebem o valor 1 em todas as valorações possíveis).

DEFINIÇÃO:

Uma fórmula α é uma **tautologia** se e somente se, para toda valoração $v, v(\alpha) = 1$.

Uma fórmula α é uma **contradição** se e somente se, para toda valoração $v, v(\alpha) = 0$.

Uma fórmula α é uma **contingência** se e somente se não for tautologia, nem for contradição.

Os dois exemplos de tabelas acima provam que as fórmulas $\neg\neg\neg A$ e $\neg(A \rightarrow \neg B)$ são contingentes. Vejamos, agora, dois exemplos bem simples, um de tautologia e outro de contradição, devidamente provados pelas respectivas tabelas:

A	$\neg A$	$\neg\neg A$	$A \leftrightarrow \neg\neg A$
1	0	1	1
0	1	0	1

A	$\neg A$	$A \wedge \neg A$
1	0	0
0	1	0

Qual a importância de identificarmos as fórmulas tautológicas de nossa linguagem L na semântica para o CPC? As tautologias correspondem às leis (ou *verdades lógicas*) do CPC – também dizemos que são as *fórmulas válidas* do CPC. As fórmulas contraditórias são igualmente importantes porque suas negações são automaticamente tautologias – por exemplo, tente conferir como fica a tabela para a negação da fórmula contraditória acima; ou seja, construa a tabela para a fórmula $\neg(A \wedge \neg A)$.

O conjunto das tautologias é obviamente infinito; porém, listaremos abaixo, para estimular a curiosidade do leitor, apenas alguns esquemas tautológicos interessantes do CPC, cada um dos quais tem motivado discussões filosó-

ficas fascinantes, frequentemente levando à elaboração de cálculos proposicionais bem diferentes do CPC (e muitos até mesmo incompatíveis com este).

Alguns esquemas tautológicos:

Não contradição: $\neg(\alpha \wedge \neg\alpha)$

Terceiro excluído: $\alpha \vee \neg\alpha$

Dupla negação: $\alpha \leftrightarrow \neg\neg\alpha$

Comutatividade da conjunção: $(\alpha \wedge \beta) \leftrightarrow (\beta \wedge \alpha)$

Uma vez que as fórmulas tautológicas são as verdades lógicas do CPC, temos uma maneira bem simples e rigorosa de *caracterizar semanticamente* essa teoria lógica: consiste exatamente no conjunto de todas as tautologias. A outra maneira de construir o CPC e obter as mesmas verdades lógicas é pela via sintática, como veremos na próxima subseção, para depois discutirmos brevemente algumas relações importantes entre essas duas perspectivas (semântica e sintática) para o CPC.

Antes disso, mencionaremos mais uma noção semântica importante, devido a sua relação com o conteúdo da subseção 2.2, especificamente, as noções de consequência lógica e de argumento válido. Naquela subseção, fornecemos definições informais para ambas as noções. Aqui, apresentaremos definições mais precisas usando as noções semânticas introduzidas.

DEFINIÇÃO: Uma valoração v é **modelo** de um conjunto Γ de fórmulas (em símbolos, $v \models \Gamma$) se e somente se, para toda fórmula $\gamma \in \Gamma$, $v(\gamma) = 1$

Ou seja, um modelo de um conjunto de fórmulas é uma valoração que torne todas as fórmulas daquele conjunto verdadeiras ao mesmo tempo.

Sendo assim, como avaliar se um argumento é válido (de acordo com a definição da subseção 2.2)? A estrutura de um argumento pode ser considerada como um conjunto (finito) de fórmulas representando suas premissas e uma

dada fórmula representando sua conclusão. Assim, a definição anteriormente fornecida para consequência lógica pode ser assim reformulada:

DEFINIÇÃO: Seja Γ um conjunto de fórmulas, e α uma fórmula qualquer. Dizemos que α é **consequência lógica (semântica)** de Γ (em símbolos, $\Gamma \models \alpha$) se e somente se, para toda valoração v tal que $v \models \Gamma$, $v(\alpha) = 1$.

Ou seja, toda valoração que seja modelo de Γ também deve ser modelo de α . (Recomendamos comparar com a definição mais informal, fornecida antes, para perceber a equivalência entre elas!)

Para averiguar formalmente essa relação de consequência lógica, podemos empregar o método das tabelas de verdade, explicado acima. Basta construir uma tabela que compare, para cada valoração possível, os valores de cada premissa em Γ e o valor da conclusão α . Por exemplo, retomemos o argumento seguinte, já analisado anteriormente:

P₁: Ou João está na universidade, ou está em casa.

P₂: João não está em casa.

C: Por conseguinte, João está na universidade.

Uma simbolização bem intuitiva para esse argumento pode usar a seguinte convenção para fazer corresponder as seguintes fórmulas atômicas de **L** com sentenças atômicas do português:

A: João está na universidade.

B: João está em casa.

Deixando o argumento com a seguinte simbolização em **L**: $A \vee B$, $\neg B \therefore A$

Para testar se o conjunto $\{A \vee B, \neg B\}$ de premissas implica logicamente (ou seja, tem como consequência lógica) a fórmula A – ou seja, se $\{A \vee B, \neg B\} \models A$

$B, \neg B\} \models A$ – construímos uma tabela de verdade como essa abaixo e conferimos se todas as valorações que tornam as fórmulas do conjunto $\{A \vee B, \neg B\}$ verdadeiras ao mesmo tempo sempre tornam simultaneamente a conclusão A verdadeira. Como isso de fato é o caso (a única linha em que ambas as premissas são verdadeiras é a terceira, e a conclusão também é verdadeira na mesma linha), temos que a conclusão é uma consequência lógica daquele conjunto de premissas.

A	B	$A \vee B$	$\neg B$	A
1	1	1	0	1
0	1	1	0	0
1	0	1	1	1
0	0	0	1	0

Para encerrar, consideremos outro argumento já usado anteriormente, seguido de uma convenção para a simbolização de suas sentenças atômicas como fórmulas atômicas de L :

P_1 : Se João está tocando bateria, então ele está no quarto.

P_2 : João está no quarto.

C: Portanto, João está tocando bateria.

A : João está tocando bateria.

B : João está no quarto.

O argumento fica, então, com a seguinte forma: $A \rightarrow B, B \therefore A$. É fácil perceber, pela tabela abaixo, que $A \rightarrow B, B$ não é consequência lógica do conjunto $\{A \rightarrow B, B\}$ de premissas. Embora na primeira linha, a conclusão coincidiu de ser verdadeira ao mesmo tempo em que as premissas, na segunda linha a verdade das premissas não garantiu a verdade da conclusão. Logo, não acontece que $\{A \rightarrow B, B\} \models A$.

A	B	$A \rightarrow B$	B	A
1	1	1	1	1
0	1	1	1	0
1	0	0	0	1
0	0	1	0	0

(Por simplicidade, seguindo o uso mais corrente, no restante de nossa exposição, omitiremos as chaves na indicação do conjunto de premissas de um argumento logicamente válido. Por exemplo, em vez de $\{A \vee B, \neg B\} \models A$, escreveremos apenas $A \vee B, \neg B \models A$.)

Outra convenção importante, acompanhando a notação mais usual. Usaremos o mesmo símbolo ' \models ' para indicar que uma fórmula qualquer α é válida ou tautológica, da seguinte maneira: $\models \alpha$. Essa expressão, na prática, abrevia " $\emptyset \models \alpha$ ", o que significa dizer que α é consequência lógica de todo e qualquer conjunto de premissas (até do conjunto vazio); ou seja, α vale em qualquer caso.

3.3. Uma sintaxe para o CPC e a noção sintática de consequência lógica

Nesta subseção, vamos expor a contraparte sintática do CPC de maneira bem geral. Dentre as noções sintáticas, apresentaremos a noção de consequência lógica, sem recorrer a algum tipo de interpretação, como fizemos na subseção anterior através da noção de valoração. Para tanto, recorreremos à noção de sistema formal.

DEFINIÇÃO: Um **sistema formal** é um par ordenado $\langle \mathcal{L}, \wp \rangle$, onde \mathcal{L} é uma linguagem formal e \wp é um conjunto de postulados.

Uma *linguagem formal*, como vimos na subseção 3.1, envolve um conjunto de símbolos (o alfabeto) e um conjunto de regras de formação (a gramática). *Postulados* aqui são entendidos como princípios que são assumidos sem demonstração. Um conjunto de postulados pode conter axiomas ou regras de inferências. *Axiomas* são fórmulas que são usadas como suposições iniciais

(sem demonstração) das quais outras fórmulas são derivadas. *Regras de inferências* são relações entre fórmulas que são assumidas (sem demonstração) que nos dizem como gerar (derivar) novas fórmulas a partir de axiomas ou outras fórmulas já derivadas.

Essa noção abstrata de sistema formal vai ficar mais clara a partir da apresentação de um sistema formal muito simples, usando o método da dedução natural, que iremos expor aqui para o CPC. A linguagem \mathcal{L} do sistema formal que apresentaremos será a própria linguagem L que já conhecemos, definida na subseção 3.1. O conjunto de postulados do nosso sistema formal terá apenas regras de inferências; seu conjunto de axiomas é vazio.⁴ Mas, antes de procedermos com nosso sistema de dedução natural, poderíamos questionar a utilidade de um sistema formal de prova. Por que adotar um sistema formal para verificar se uma dada conclusão é consequência lógica de um conjunto de premissas, uma vez que temos um método efetivo como o de tabelas de verdade? Bom, há algumas razões para se adotar um sistema formal. Aqui colocaremos uma razão prática. Suponha um argumento como o seguinte:

$$A \vee B, B \rightarrow C, C \rightarrow D, \neg A, D \rightarrow E \vDash (E \wedge B) \vee F$$

Como temos seis variáveis proposicionais nesse argumento, teríamos de construir uma tabela de verdade com 64 linhas para testar sua validade lógica, se formos recorrer ao método semântico de tabelas de verdade. Um sistema formal frequentemente fornece, em situações assim, uma maneira mais compacta para testar se a conclusão é consequência das premissas. A ideia é aplicar as regras de inferência postuladas ao conjunto de premissas, obtendo conclusões intermediárias, até chegar na conclusão do argumento. Quando alcançamos a fórmula que está na conclusão, a partir desse procedimento, dizemos que deduzimos (demonstramos) a fórmula da conclusão a partir das premissas. Intuitivamente, as regras de inferência são argumentos mais curtos que nos auxiliam a alcançar a conclusão de um argumento mais longo. O caminho é constituído por fórmulas que aparecem em uma sequência, as quais podem ser premissas ou fórmulas que já foram obtidas de fórmulas anteriores, por aplicação de alguma

⁴Há apresentações sintáticas do CPC que são feitas através de sistemas que assumem axiomas (ver MENDELSON, 2010, seção 1.4.)

regra de inferência (ou ainda axiomas, caso o sistema seja axiomático). Esse caminho é chamado de dedução (que será apresentado de maneira mais rigorosa ainda nesta subseção).

Tome, por exemplo, $A \vee B, \neg B \vDash A$. Conferimos, na última seção, que esse argumento curto é válido. Poderíamos postular a forma desse argumento como uma regra de inferência esquematicamente da seguinte maneira:

$$\frac{\alpha \vee \beta \quad \neg \alpha}{\beta}$$

O que essa regra de inferência diz é o seguinte: se há $\alpha \vee \beta$ em uma linha da demonstração e na outra linha há $\neg \alpha$, estamos autorizados a deduzir β . O traço horizontal indica que a fórmula abaixo do traço pode ser derivada (deduzida), através da regra em questão, a partir das fórmulas que estão acima do traço. Note que aqui não estamos falando do significado das fórmulas envolvidas. Neste contexto sintático, a derivação consiste basicamente em manipulações de símbolos (no caso, as fórmulas).

Com isso, do ponto de vista formal, nada nos impede de postularmos falácias como regras de inferências do sistema. Todavia, o sistema resultante não seria útil se quisermos que ele verifique exatamente aqueles argumentos que resultaram válidos. Almejamos um sistema que seja correto e que, portanto, capture os argumentos válidos (semanticamente). Para tanto, devemos escolher regras que são preservadoras de verdade: regras que, quando aplicadas às fórmulas verdadeiras, resultam sempre em fórmulas verdadeiras.

Feita essas considerações iniciais, vejamos, sem muitos detalhes, as regras do sistema de dedução natural apresentado aqui. Como **L** tem cinco operadores lógicos, teremos dez regras de inferências, uma regra de introdução e uma de eliminação para cada operador. Especificamente, teremos um par de regras para cada um dos seguintes operadores: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$. (Em cada caso no quadro abaixo, indicamos uma notação curta para referir-se à regra definida, bem como um nome em português para a mesma regra.)

+\wedge (Introdução da conjunção)		-\wedge (Eliminação da conjunção)	
$\frac{\alpha}{\beta}$ <hr style="width: 50%; margin: auto;"/> $\alpha \wedge \beta$		$\frac{\alpha \wedge \beta}{\alpha}$	$\frac{\alpha \wedge \beta}{\beta}$
+\vee (Introdução da disjunção)		-\vee (Eliminação da disjunção)	
$\frac{\alpha}{\alpha \vee \beta}$	$\frac{\alpha}{\beta \vee \alpha}$	$\frac{\alpha \vee \beta}{\neg \alpha}$ <hr style="width: 50%; margin: auto;"/> β	$\frac{\alpha \vee \beta}{\neg \beta}$ <hr style="width: 50%; margin: auto;"/> α
+\leftrightarrow (Introdução da bi-implicação)		-\leftrightarrow (Eliminação da bi-implicação)	
$\frac{\alpha \rightarrow \beta}{\beta \rightarrow \alpha}$ <hr style="width: 50%; margin: auto;"/> $\alpha \leftrightarrow \beta$		$\frac{\alpha \leftrightarrow \beta}{\alpha \rightarrow \beta}$	$\frac{\alpha \leftrightarrow \beta}{\beta \rightarrow \alpha}$

$+ \rightarrow$ (Introdução da implicação)	$- \rightarrow$ (Eliminação da implicação)
$\begin{array}{l} \alpha \\ \\ \cdot \\ \cdot \\ \cdot \\ \beta \\ \hline \alpha \rightarrow \beta \end{array}$	$\frac{\alpha \rightarrow \beta \quad \alpha}{\beta}$
$+ \neg$ (Introdução da negação)	$- \neg$ (Eliminação da negação)
$\begin{array}{l} \alpha \\ \\ \cdot \\ \cdot \\ \cdot \\ \beta \wedge \neg \beta \\ \hline \neg \alpha \end{array}$	$\frac{\neg \neg \alpha}{\alpha}$

Como deve ter ficado evidente, para denotar as regras de introdução de um operador, estamos usando o sinal '+' seguido do operador e, analogamente, para denotar as regras de eliminação de um operador, estamos usando o sinal '-'. Assim, por exemplo, '- V' significa eliminação da disjunção. A propósito, as regras da tabela acima são conhecidas por outros nomes. O esquema inferencial que usamos para eliminação da disjunção no sistema em questão é conhecido como silogismo disjuntivo. Em vez de silogismo disjuntivo, poderíamos ter usado outro esquema inferencial para eliminação da disjunção. Para desempenhar o

papel de eliminação da disjunção, alguns sistemas de dedução natural trazem a regra conhecida como *prova por caso*:

$$\frac{\alpha \vee \beta \quad \alpha \rightarrow \gamma \quad \beta \rightarrow \gamma}{\gamma}$$

Bom, mas como não há a regra de prova por casos dentre as regras do sistema de dedução natural apresentado acima, essa regra não ficaria faltando? Isto é, será que há algum argumento que é válido, mas que não poderia ser provado pela falta da regra de prova por casos? A resposta é não, pois essa regra pode ser derivada (demostrada ou deduzida), a partir das regras de inferências primitivas, apresentadas no quadro acima. Para demonstrar isso, podemos tomar um argumento que instancia o esquema metalinguístico da prova por caso em que derivamos C a partir do seguinte conjunto de premissas: $A \vee B, A \rightarrow C, B \rightarrow C$. (Deixaremos essa demonstração para o leitor fazer depois de terminar de ler esta subseção).

Vamos agora definir com um pouco mais de rigor as noções de dedução e consequência lógica (sintática) que estamos tratando aqui:

DEFINIÇÃO: Sejam Γ um conjunto qualquer de fórmulas e α uma fórmula. Uma **dedução** de α a partir de Γ é uma sequência finita $\delta_1, \dots, \delta_n$ de fórmulas, tal que $\delta_n = \alpha$ e cada $\delta_i, 1 \leq i \leq n$, é uma fórmula que pertence a Γ ou foi obtida a partir de fórmulas que aparecem antes na sequência, por meio da aplicação de alguma regra de inferência.

Ou seja, uma dedução de uma determinada fórmula α a partir de um conjunto Γ de fórmulas é uma sequência finita de fórmula $\delta_1, \dots, \delta_n$ (aquilo que estávamos chamando informalmente por caminho) em que o último elemento da sequência, δ_n , é a própria fórmula α . Cada fórmula, δ_i , da sequência ou pertence ao conjunto Γ de premissas ou foi obtida de fórmula(s) que aparecia(m) antes, por meio da aplicação de uma regra de inferência. Tipicamente, as deduções são

numeradas, contendo em cada linha, uma fórmula e mais uma justificativa para a fórmula estar presente na dedução, como veremos abaixo através de um exemplo.

Com base na noção de dedução definida acima, podemos agora apresentar uma definição mais rigorosa da noção de consequência lógica (sintática):

DEFINIÇÃO: Seja Γ um conjunto de fórmulas e α uma fórmula. Dizemos que α é **consequência lógica (sintática)** de Γ (em símbolos $\Gamma \vdash \alpha$) se há uma dedução de α a partir de Γ .

Para deixar essas noções menos abstratas, tomaremos um exemplo de argumento e verificaremos (de maneira sintática) se sua conclusão é consequência lógica das premissas. Vamos aproveitar e já conferir o argumento que colocamos no início desta subseção, quando dissemos que, através do método de dedução natural que estamos vendo, teríamos uma maneira mais compacta que o método das tabelas de verdade para verificar a validade de alguns argumentos. O argumento usado como exemplo tinha como conclusão a fórmula $(E \wedge B) \vee F$ e o seguinte conjunto Γ de premissas: $\{A \vee B, B \rightarrow C, C \rightarrow D, \neg A, D \rightarrow E\}$. O procedimento será dispor no início da sequência as premissas (justificando com o 'p' de premissa) e tentar alcançar a conclusão através das regras de inferências do sistema de dedução natural. Todas as fórmulas que aparecerem na sequência da dedução devem estar justificadas (evidenciando a linha que foi utilizada e a regra que permitiu a derivação).

1. $A \vee Bp$
2. $B \rightarrow Cp$
3. $C \rightarrow Dp$
4. $\neg Ap$
5. $D \rightarrow Ep$
6. B 1,4 / \vee
7. C 2,6 / \rightarrow
8. D 3,7 / \rightarrow
9. E 5,8 / \rightarrow
10. $E \wedge B$ 6,9 / \wedge
11. $(E \wedge B) \vee F$ 10 / \vee

Como a fórmula $(E \wedge B) \vee F$ que está na conclusão do argumento foi deduzida, conforme a definição de dedução, partindo do conjunto de premissas $\{A \vee B, B \rightarrow C, C \rightarrow D, \neg A, D \rightarrow E\}$, temos que $(E \wedge B) \vee F$ é consequência lógica sintática de $\{A \vee B, B \rightarrow C, C \rightarrow D, \neg A, D \rightarrow E\}$, conforme a definição de consequência lógica sintática. Assim, em símbolos, temos $\{A \vee B, B \rightarrow C, C \rightarrow D, \neg A, D \rightarrow E\} \vdash (E \wedge B) \vee F$.

É interessante destacar que as regras que foram colocadas para introdução da negação e para introdução da implicação na tabela de regras de inferências acima, são regras que envolvem hipóteses (regras hipotéticas). No sistema que estamos apresentando aqui, a regra usada para introdução da negação é conhecida como *redução ao absurdo* (RAA) e a regra usada para introdução da implicação é conhecida como *regra de prova condicional* (RPC). Como exemplo, vejamos o funcionamento da regra de redução ao absurdo. Para tanto vamos conferir um argumento simples, inspirado na citação de Descartes, exposta no início deste capítulo. Imaginamos o seguinte argumento:

P₁: Se Deus é enganador, então ele está sujeito à carência

P₂: Deus não está sujeito a carência

C: Logo, Deus não é enganador.

Simbolizando o argumento para a linguagem proposicional L , temos a fórmula $\neg E$ como conclusão do argumento com as seguintes premissas: $E \rightarrow C$, $\neg C$. Para verificar se a conclusão é consequência lógica sintática das premissas, considere a seguinte dedução:

1.	$E \rightarrow C$	p
2.	$\neg C$	p
3.	E	h
4.	C	$1, 3 / \rightarrow$
5.	$C \wedge \neg C$	$2, 4 / + \wedge$
6.	$\neg E$	$3 - 5 / + \neg$

Temos, portanto, $\{E \rightarrow C, \neg C\} \vdash \neg E$.

Apenas alguns comentários rápidos sobre a dedução acima: após as premissas, foi colocada a fórmula E como hipótese h (para denotar a hipótese). Grosso modo, a ideia da regra de RAA, usada como introdução da negação, é de que, se a hipótese levar à contradição (uma fórmula da forma $\beta \wedge \neg\beta$), estamos autorizados, através da regra em questão, a fechar a hipótese (denotado pela barra vertical na demonstração) e inferir a negação da fórmula que aparece na hipótese.⁵ O argumento em questão é uma instância de uma forma de argumento conhecida como *modus tollens*: $\alpha \rightarrow \beta, \beta \vdash \neg\alpha$.⁶ Uma vez demonstrada, essa forma de argumento pode entrar como uma regra de inferência fazendo com que as demonstrações fiquem mais curtas e, com isso, podemos ter várias regras desse tipo no sistema de dedução natural. Para dar mais um exemplo, a regra por casos, mencionada acima, também pode entrar como uma

⁵ Considerando o que foi dito na introdução, que o presente texto não pretende desempenhar o papel de um curso básico de lógica, não vamos tratar de expor o funcionamento de todas as regras de inferência. O objetivo aqui é apresentar de maneira sucinta o CPC como um sistema de lógica. Idealmente esse texto pode servir como uma motivação para o leitor proceder com um estudo mais detalhado através dos manuais de lógica indicados no final deste texto.

⁶ Confira, no capítulo 'Filosofia da Ciência' incluído neste volume, uma aplicação do *Modus Tollens* na filosofia de Karl Popper.

regra do sistema de dedução natural em questão. O ponto é que, no sistema que estamos apresentando, tais regras não são primitivas, uma vez que não foram postuladas (assumidas sem demonstração) no sistema. Elas são *regras derivadas*, são obtidas a partir das regras primitivas, e, sendo assim, elas são em certo sentido redundantes: tudo que podemos demonstrar com elas pode ser demonstrado sem elas (i.e., só com as regras primitivas). Ao encurtarem as demonstrações, podemos dizer que elas funcionam como uma espécie de atalho no caminho (dedução) que devemos percorrer.

Para encerrar esta breve seção sobre o CPC, vamos falar rapidamente da noção de um caso particular de consequência lógica sintática. Quando temos uma fórmula que é consequência lógica do conjunto vazio de premissas, chamamos essa fórmula de *teorema*.

DEFINIÇÃO: Uma fórmula α é um **teorema** se há uma dedução de α a partir do conjunto vazio de premissas. I.e., α é um teorema sse $\emptyset \vdash \alpha$, o que abreviamos por $\vdash \alpha$.

Uma vez que, à primeira vista, pode parecer um pouco estranho uma fórmula ser consequência lógica do conjunto vazio, vamos colocar um exemplo simples de teorema do CPC: $\neg\neg A \leftrightarrow A$. Se, até então, as deduções começavam com as premissas do argumento, como fazer para deduzir uma fórmula a partir do conjunto vazio de premissas? Para tanto, no sistema de dedução natural aqui apresentado, temos que iniciar com regras hipotéticas. Neste caso específico, com a fórmula, $\neg\neg A \leftrightarrow A$, vamos fazer uso da regra de prova condicional (a introdução da conjunção no sistema de dedução natural aqui apresentado) e aproveitar para ver o funcionamento desta regra:

1.	$\neg\neg A$	h
2.	A	1 / \neg
3.	$\neg\neg A \rightarrow A$	1 - 2 / $+$ \rightarrow
4.	A	h
5.	$\neg A$	h
6.	$A \wedge \neg A$	4, 5 / $+$ \wedge
7.	$\neg\neg A$	5 - 6 / $+$ \neg
8.	$A \rightarrow \neg\neg A$	4 - 7 / $+$ \rightarrow
9.	$\neg\neg A \leftrightarrow A$	3, 8 / $+$ \leftrightarrow

Como $\neg\neg A \leftrightarrow A$ é demonstrada a partir do conjunto vazio de premissas, temos que ela é um teorema do CPC. Uma vez provado esse teorema, ele pode servir como um esquema de teorema, $\neg\neg\alpha \leftrightarrow \alpha$, em que podemos substituir uniformemente a metavariable α por fórmulas da linguagem obtendo, assim, uma infinidade de teoremas como, por exemplo, $\neg\neg B \leftrightarrow B$ ou $(A \vee B) \leftrightarrow \neg\neg(A \vee B)$, etc. É interessante notar que os teoremas são contrapartes sintáticas da noção de fórmula válida (tautologia). As fórmulas válidas do CPC (as tautologias) mencionadas acima podem ser demonstradas como teoremas no CPC. Tendo isso em vista, como exercício, sugerimos ao leitor demonstrar uma instância de cada esquema tautológico que foi exposto na subseção 3.2.

4. Considerações finais: metalógica, lógicas e filosofia da lógica

Neste texto introdutório, procuramos fazer uma apresentação muito geral sobre a Lógica, vista como análise de argumentos, e algumas questões relacionadas a essa ciência. Nosso objetivo com este breve capítulo, como dissemos, não foi fornecer um curso de Lógica, como é feito em manuais especializados. Existem excelentes cursos de Lógica em manuais específicos para isso (citamos e recomendamos alguns desses manuais na lista de sugestões de leitura abaixo). Nosso objetivo, bem mais modesto, foi ao longo do texto colocar o leitor em contato com alguns conceitos lógicos importantes, apresentando, sem muito detalhes, uma teoria lógica muito simples; o Cálculo Proposicional Clássico (CPC). Para encerrar, agora faremos alguns comentários muito breves sobre

metalógica, lógicas e filosofia da lógica a fim ampliar um pouco mais o horizonte da Lógica, oferecendo algumas sugestões para o leitor interessado.

Vimos algumas noções gerais de argumentos e evidenciamos que, como um conjunto de teorias sobre a relação de consequência, a Lógica pode lidar com argumentos. Neste capítulo, exploramos a relação de consequência do CPC em uma perspectiva semântica, através do método das tabelas de verdade, e em uma perspectiva sintática, através de um sistema formal de dedução natural. Vimos que a noção teórica de consequência lógica semântica, em certo aspecto, captura a ideia intuitiva, ou pré-teórica, de validade como preservação de verdade. Já na parte sintática, como vimos, embora um sistema formal possa ser visto de um ponto de vista estritamente formal, sem nenhuma semântica pretendida, o sistema de dedução natural tratado aqui captura a ideia de preservação de verdade.

Podemos perceber que as noções semântica e sintática de consequência lógica do CPC, mesmo sendo distintas, podem apresentar conexões interessantes. Particularmente, quando o sistema formal apresenta propriedades meta-teóricas interessantes, como a correção e a completude. Grosso modo, quando dizemos que um sistema formal é correto e completo, as duas relações de consequência lógica, sintática e semântica, coincidem em um certo sentido (extensional).

Um sistema formal é *correto* sempre que, se uma fórmula α for consequência lógica sintática de um conjunto Γ de fórmulas (em símbolos, $\Gamma \vdash \alpha$), α será também consequência lógica semântica de Γ (em símbolos, $\Gamma \models \alpha$). Um sistema formal é *completo* sempre que uma fórmula α for consequência lógica semântica de um conjunto Γ de fórmulas (em símbolos, $\Gamma \models \alpha$), α será consequência lógica sintática de Γ (em símbolos, $\Gamma \vdash \alpha$). Colocando a correção e a completude juntas, temos o seguinte: $\Gamma \vdash \alpha$ se e somente se $\Gamma \models \alpha$.

Um caso particular da correção e da completude é quando o conjunto Γ é vazio. Assim, em um sistema formal correto e completo, uma fórmula α é teorema (em símbolos, $\vdash \alpha$), se e somente se α é válida ($\models \alpha$). Um sistema formal correto demonstra *somente* as fórmulas válidas, enquanto um sistema formal completo demonstra *todas* as fórmulas válidas. Especificamente, o sistema de dedução natural apresentado neste capítulo é correto e completo. Não cabe neste pequeno texto apresentar a prova de correção e completude para um

sistema do CPC (para isso, ver MENDELSON, 2010 e FEITOSA & PAULOVICH, 2005).

Em um sistema formal, como vimos, podemos provar, dentre outras coisas, que certas sentenças são consequência lógica de certos conjuntos. Mas, quando nos ocupamos de *propriedades gerais* sobre o sistema, como, por exemplo, se ele é correto ou completo, estamos em um campo da Lógica conhecido como *metalógica*, e as ditas propriedades e resultados são também chamados de *metalógicos*. Há várias outras propriedades e resultados importantes na metalógica (como decidibilidade, consistência, etc.); porém, os exemplos acima devem bastar como ilustração de resultados metalógicos.

Mencionamos, em diversos momentos, a existência de várias lógicas (i.e., teorias da relação de consequência). O CPC visto brevemente aqui pode ser expandido para uma teoria mais forte, o Cálculo Quantificacional Clássico (CQC). O CPC é entendido como um subsistema do CQC no sentido de que tudo que pode ser expresso e provado no CPC pode ser expresso e provado no CQC. A teoria lógica do CQC tem mais princípios e sua linguagem tem um poder expressivo maior que a do CPC. Tradicionalmente o CQC, que inclui o CPC, é chamado de lógica elementar (para uma apresentação bastante didática do CQC, indicamos MORTARI, 2016).

Todavia, além das teorias lógicas clássicas, há teorias (ou lógicas) não clássicas. Podemos classificar as lógicas não clássicas em dois grupos: as lógicas que estendem a lógica clássica, adicionando um vocabulário distinto do vocabulário clássico (como operadores que não são funções de verdade, por exemplo), chamadas de lógicas ampliadas, e lógicas que derogam algum princípio importante da lógica clássica (como o princípio da bivalência, terceiro excluído, não contradição), que são chamadas de lógicas heterodoxas ou alternativas. É importante destacar que essa classificação de lógicas não clássicas não está isenta de discussão. Há vários entendimentos e classificações sobre lógicas não clássicas que não serão mencionados neste texto (para tanto, indicamos BURGESS, 2009; DA COSTA, 2008; HAACK, 2002).

Um exemplo de lógicas não clássicas ampliadas são os numerosos sistemas de lógicas modais que incluem a lógica clássica. Entretanto, há inúmeros outros sistemas modais que são ampliados a partir de lógicas não clássicas heterodoxas tomadas como base (daí percebemos que a classificação entre lógica não clássicas ampliativas e heterodoxas tem algumas ressalvas). Grosso

modo, lógicas modais são lógicas que contêm operadores modais (que não são funções de verdade) que podem ter interpretações aléticas, ('é necessário que', 'é possível que'), temporais ('será sempre o caso que', 'foi o caso que'), deônticas ('é obrigatório', 'é permitido'). Há vários sistemas de lógicas modais normais (como K, T, D, B, S4, S5) e não normais (como S1, S2, S3, E, E2, E3). (Para saber mais sobre lógica modal, veja SIDER, 2010, caps. 6 e 7 e MORTARI, 2016, cap. 18).

Como exemplo de lógicas alternativas, cabe mencionar rapidamente três exemplos: lógicas polivalentes, lógicas intuicionistas e lógicas paraconsistentes. Como vimos neste capítulo, as fórmulas do CPC são bipartidas; ou são verdadeiras ou são falsas. Para o princípio da bivalência, o importante é a bipartição do conjunto das fórmulas, independentemente da natureza dos valores de verdade em questão. Embora falemos de valores de verdade, a lógica clássica não está comprometida com a interpretação de seus valores como sendo o verdadeiro ou o falso. Podemos, por exemplo, dar uma interpretação para os valores do CPC através de circuitos eletrônicos, preservando ainda o princípio clássico da bivalência (ver FEITOSA & PAULOVICH, 2005, cap. 4). Todavia, há lógicas que derogam o princípio da bivalência. Essas são chamadas de lógicas polivalentes ou multivaloradas. No geral, lógicas polivalentes são lógicas com n valores de verdade, com $n > 2$. Como exemplo de lógicas polivalentes, podemos citar a lógica L_3 de Łukasiewicz, que assume três valores de verdade: o verdadeiro, o falso e o indeterminado. (Para saber mais sobre lógicas polivalentes, veja HAACK, 2002, cap. 11; PRIEST, 2008, cap. 7).

A lógica clássica também é regida pelo princípio do terceiro excluído que pode ser enunciado da seguinte maneira: dada uma proposição, α , e sua negação, $\neg\alpha$, ao menos uma é verdadeira. Como vimos na subseção 3.2, este princípio pode ser formulado sintaticamente na lógica proposicional clássica da seguinte forma: $\alpha \vee \neg\alpha$. Todavia, há lógicas, como as lógicas intuicionistas, que derogam o princípio clássico do terceiro excluído. Em linhas gerais, para a perspectiva intuicionista, o que torna uma fórmula verdadeira é uma prova construtiva dessa fórmula. Assim, uma regra de inferência como a de redução ao absurdo, que vimos na subseção 3.3, não vale na lógica intuicionista. Uma das motivações vem do fato que há problemas na matemática que, até então, não foram resolvidos. Tome, por exemplo, a conjectura de Goldbach, que diz que todo número inteiro e par maior do que dois é igual à soma de dois primos. Va-

mos chamar a proposição que expressa essa conjectura de G . Até hoje os matemáticos não encontraram nenhuma prova construtiva para G , nem uma a prova para $\neg G$. Assim, de um ponto de vista intuicionista, não estamos autorizados a afirmar $G \vee \neg G$, e nem $\alpha \vee \neg \alpha$ em geral. (Para apresentações de lógica intuicionista, veja PRIEST, 2008, cap. 6 e RODRIGUES, 2011, pp. 70 – 79).

Na lógica clássica vale também o chamado princípio da explosão, que, intuitivamente, diz que de uma contradição pode ser inferida qualquer sentença. Podemos formular sintaticamente tal princípio da seguinte maneira: $\alpha \wedge \neg \alpha \vdash \beta$. É fácil mostrar que esse princípio vale na lógica clássica (como exercício, deixamos para o leitor a demonstração de tal princípio no sistema de dedução natural apresentado em 3.3). Lógicas que violam o princípio clássico da explosão são chamadas de lógicas paraconsistentes. Há várias lógicas paraconsistentes como, por exemplo, as lógicas Cn de da Costa e a lógica do paradoxo (LP) de Priest. Vamos falar um pouquinho sobre essa última. Além de paraconsistente, LP pode ser vista como uma lógica de três valores, tendo sentenças que são verdadeiras, sentenças falsas e sentenças verdadeiras e falsas. A motivação para assumir sentenças que são verdadeiras e falsas vem da doutrina conhecida como dialeteísmo, a visão de que há contradições verdadeiras. Um exemplo de contradição verdadeira dada pelos dialeteístas, como Priest, é o paradoxo do mentiroso que pode ser alcançado através de uma sentença que afirma a sua própria falsidade (conhecida como sentença do mentiroso): esta sentença é falsa. A questão é: qual é o valor de verdade da sentença do mentiroso? Se, por um lado, ela for verdadeira, ela é falsa. Se, por outro lado, ela for falsa, ela é verdadeira. Para os dialeteístas, como Priest, o paradoxo do mentiroso mostra que há contradições, $\alpha \wedge \neg \alpha$, verdadeiras. Como o dialeteísmo é a visão de que algumas contradições são verdadeiras, e não que todas as sentenças são verdadeiras, o princípio de explosão não pode ser válido. (Para uma apresentação da lógica paraconsistente LP e algumas noções sobre o dialeteísmo, veja PRIEST, 2008, cap. 7). Cabe mencionar aqui que, embora o dialeteísmo (uma visão sobre a verdade) leve à paraconsistência (uma teoria sobre a relação de consequência), a paraconsistência não está comprometida com o dialeteísmo. As lógicas de Da Costa, por exemplo, não estão comprometidas com a verdade de contradições (veja Da Costa, 2008). Há outras interpretações da paraconsistência que rejeitam a visão dialeteísta, de que contradições podem ser verdadeiras, e inter-

pretam a contradição em termos de evidência, e não em termos de verdade (Cf. CARNIELLI & RODRIGUES, 2017).

A questão sobre a interpretação de uma lógica, como a paraconsistente, está no campo da filosofia da lógica. A filosofia da lógica pode ser entendida como uma área da filosofia que, grosso modo, discute questões levantadas pela lógica. Uma questão interessante que levantamos aqui, no começo deste texto, é a questão sobre a natureza da Lógica. O que é Lógica? Responder a essa pergunta é um trabalho de reflexão filosófica acerca da Lógica (ver COHNITZ & ESTRADA GONZALES, 2019, cap. 1).

A filosofia da lógica também trata de paradoxos como o paradoxo do mentiroso, que surge a partir da noção de verdade (ver HAACK, 2002, cap. 8; CARDOSO, 2018) e paradoxos sobre a implicação material (ver READ, 2016, cap. 3). Os paradoxos sobre a noção de verdade também contribuem para a discussão filosófica sobre a natureza da verdade (ver HAACK, cap. 7) dando espaço para o surgimento de teorias formais da verdade (ver BEALL, GLANZBERG, RIPLEY, 2018). Paradoxos e outros fenômenos que surgem na Lógica também colocam em questão se a Lógica pode ser revisada (COHNITZ & ESTRADA GONZALES, 2019, cap. 6 e PRIEST, 2014).

Outra reflexão filosófica interessante no campo da filosofia da lógica é sobre se há apenas uma lógica correta que corretamente codifica a relação de consequência (monismo) ou há uma pluralidade de lógicas que codificam corretamente a relação de consequência (pluralismo) (ver COHNITZ & ESTRADA GONZALES, 2019, cap. 6; HAACK, 2002, cap. 12). Além dessas questões apenas mencionadas aqui, há outras reflexões filosóficas sobre a Lógica. Frente aos vários problemas filosóficos sobre a Lógica, percebe-se que essa ciência não é importante apenas como um conjunto de teorias sobre a relação de consequência que podem ter distintas aplicações, como em análise de argumentos, mas que ela por si só é uma disciplina viva e rica do ponto de vista filosófico.

Esperamos que estas considerações preliminares sobre a Lógica tenham oferecido um panorama de algumas características dessa importante área de estudos, tradicionalmente estudada dentro da Filosofia, por suas múltiplas aplicações (metafísicas, epistemológicas, semânticas, etc.), bem como pelas provocações filosóficas que oferece.

Questões para revisão:

1. Explique o que são argumentos e como algumas expressões gramaticais podem servir de indicadores para os elementos de sua estrutura.
2. Explique a diferença entre argumentos válidos e argumentos corretos.
3. Compare e diferencie entre deduções, induções e abduções. Elabore exemplos novos para cada um desses tipos de argumentos.
4. Explique por que as falácias não são aceitáveis numa discussão racional, tendo em vista que outros tipos de argumentos inválidos podem ser legítimos nesse contexto.
5. Considerando-se que são possíveis infinitas valorações para as variáveis proposicionais de uma linguagem formal, por que não precisamos de infinitas linhas na tabela de verdade de uma fórmula dessa linguagem?
6. Compare e diferencie as noções de consequência lógica semântica e consequência lógica sintática.
7. Compare e diferencie as noções de tautologia (ou fórmula válida) e teorema.
8. Explique as noções metalógicas de correção e de completude.
9. Qual a diferença entre lógicas não clássicas ampliativas e heterodoxas? Mencione exemplos, relacionando-as com alguns princípios fundamentais da lógica clássica.
10. Qual a perspectiva investigativa da Filosofia da Lógica e como ela se diferencia da metalógica e da própria Lógica?

Questões para discussão:

1. Se nem todo argumento válido é bom para uma discussão racional, qual seria a importância ou relevância do estudo da Lógica, enquanto análise de argumentos?
2. É fácil mostrar que qualquer premissa tem como consequência lógica ela própria. Isso não seria um caso da falácia do círculo vicioso?
3. Com base no presente capítulo, reflita sobre a relação entre a noção pré-teórica de consequência lógica e as concepções teóricas do CPC (sintática e semântica).
4. O operador de implicação corresponde (aproximadamente) à construção condicional “se..., então,” da linguagem natural. Quais as limitações

dessa correspondência? Em outras palavras, a semântica (clássica) fornecida para fórmulas construídas com a implicação consegue capturar a relação de condição e consequência usada em nossos argumentos cotidianos? Justifique sua posição.

5. Considerando que há várias teorias lógicas (i.e., várias teorias sobre a relação de consequência lógica) incompatíveis entre si, podemos dizer que há mais de uma lógica correta? Disserte um pouco a respeito.

Sugestões de leitura em português: manuais de lógica

MORTARI, Cezar A. *Introdução à lógica*. 2. Ed. São Paulo: Editora Unesp, 2016.
COPI, I. M. *Introdução à lógica*. Trad. Álvaro Cabral. São Paulo: Mestre Jou, 1978.

FEITOSA, H. A.; PAULOVICH, L. *Um prelúdio à lógica*. São Paulo: Editora Unesp, 2005.

SMULLYAN, R. *Lógica de primeira ordem*. São Paulo: Editora Unesp, 2009.

Sugestões de leitura em português: teoria da argumentação e lógica informal

CARNIELLI, Walter. A.; EPISTEIN, Richard. L. *Pensamento crítico: o poder da lógica e da argumentação*. 4.ed. São Paulo: Editora Rideel, 2019.

WALTON, Douglas. *Lógica Informal*. Trad. Ana Lúcia R. Franco, Carlos A. Al. Salum. 2 ed. São Paulo: Ed. WMF Martins Fontes, 2012.

Sugestões de leitura em português: filosofia da lógica

DA COSTA, N. C. A. *Ensaio sobre os fundamentos da lógica*. 3. Ed. São Paulo: Hucitec/Edusp, 2008

HAACK, Susan. *Filosofia das lógicas*. Trad. Cezar Mortari; Luiz Henrique Dutra. São Paulo: Editora Unesp, 2002.

READ, S. *Repensando a lógica: uma introdução à filosofia da lógica*. Trad. Abílio Rodrigues. Belo Horizonte: Editora da UFMG, 2016.

Sugestões de leitura em português: metalógica

- BOOLOS, G.; BURGESS, J.; JEFFREY, R. *Computabilidade e Lógica*. Trad. Cezar A. Mortari. São Paulo: Editora Unesp, 2012.
- FEITOSA, H. A.; PAULOVICH, L. *Um prelúdio à lógica*. São Paulo: Editora Unesp, 2005.
- SHAPIRO, Stewart. *Lógica Clássica*. Trad. Danilo Oliveira e Rodrigo Cid. In. CID, R; DANTAS, D (orgs). *Textos selecionados de lógica* (série investigação filosófica). Pelotas: NEPFIL, 2020.

Referências

- BEALL, JC; GLANZBERG, M. RIPLEY, D. *Formal Theories of Truth*. United Kingdom: Oxford University Press, 2018.
- BOOLOS, G.; BURGESS, J.; JEFFREY, R. *Computabilidade e Lógica*. Trad. Cezar A. Mortari. São Paulo: Editora Unesp, 2012.
- BURGEES, John. *Philosophical logic*. New Jersey: Princeton University Press, 2009.
- CARDOSO, G. *O Paradoxo do Mentiroso: uma introdução*. Campinas: Editora do CLE, 2018.
- CARNIELLI, W. & RODRIGUES, A. 'An epistemic approach to paraconsistency: A logic of evidence and truth'. *Synthese*, 196, 2017.
- CARNIELLI, Walter. A.; EPISTEIN, Richard. L. *Pensamento crítico: o poder da lógica e da argumentação*. 4.ed. São Paulo: Editora Rideel, 2019.
- COHNITZ, D. & ESTRADA-GONZALES, L. *An introduction to the philosophy of logic*. New York: Cambridge University Press, 2019.
- DA COSTA, N. C. A. *Ensaio sobre os fundamentos da lógica*. São Paulo: Hucitec/Edusp, 2008.
- DESCARTES, R. *Meditações*. In. *Os Pensadores* (vol. XV). São Paulo: Abril Cultural, 1973.
- FEITOSA, H. A.; PAULOVICH, L. *Um prelúdio à lógica*. São Paulo: Editora Unesp, 2005.
- HAACK, Susan. *Filosofia das lógicas*. Trad. Cezar Mortari; Luiz Henrique Dutra. São Paulo: Editora Unesp, 2002.

- KNEALE, W; KNEALE, M. *O desenvolvimento da lógica*. Lisboa: Fundação Gulbenkian, 1980.
- MENDELSON, E. *Introduction to mathematical logic*. 5a Ed. Boca Raton-London-New York: CRC Press, 2010.
- MORTARI, Cezar A. *Introdução à lógica*. 2. Ed. São Paulo: Editora Unesp, 2016.
- PRIEST, Graham. *An introduction to non-classical logic: from if to is* (2nd ed.). Cambridge University Press, 2008.
- _____. 'Revising logic'. In P. Rush, *The Metaphysics of Logic* (pp. 211-223). Cambridge: Cambridge University Press, 2014.
- READ, S. *Repensando a lógica: uma introdução à filosofia da lógica*. Trad. Abílio Rodrigues. Belo Horizonte: Editora da UFMG, 2016.
- RODRIGUES, A. *Lógica*. São Paulo: Martins Fontes, 2011.
- SHAPIRO, Stewart. 'Lógica Clássica'. Trad. Danilo Oliveira e Rodrigo Cid. In. CID, R; DANTAS, D (orgs). *Textos selecionados de lógica* (série investigação filosófica). Pelotas: NEPFIL, 2020.
- SIDER, T. *Logic for philosophy*. United Kingdom: Oxford University Press, 2010.
- WALTON, Douglas. *Logica Informal*. Trad. Ana Lúcia R. Franco, Carlos A. Al. Salum. 2 ed. São Paulo: Ed. WMF Martins Fontes, 2012.

2

Metafísica

Rodrigo Alexandre de Figueiredo

Introdução

Procurarei aqui nessas páginas fazer uma breve introdução à metafísica. Primeiramente, abordarei o tópico sobre o que é a metafísica e sobre a sua possibilidade. A ideia é indicar alguns pressupostos históricos sobre a disciplina e algumas críticas que ela tem recebido ao longo de sua história. Na sequência, procurarei abordar quatro problemas metafísicos específicos. São eles o problema dos universais, o problema sobre os particulares concretos, o problema sobre a natureza do tempo e, por fim, o problema sobre a persistência dos particulares concretos ao longo do tempo. Todos esses problemas estão de alguma forma relacionados, como veremos no seu devido tempo. A ideia nesse segundo momento não é tentar buscar uma definição explícita de metafísica, mas entender minimamente o que é a metafísica a partir de bons exemplos de o que os metafísicos vêm fazendo ao longo da história da filosofia. E acredito que os problemas selecionados são muito relevantes para entendermos minimamente que tipo de coisa se faz em metafísica. Podemos fazer uma analogia com a matemática aqui: quando vamos aprender o que é a matemática, por vezes é mais útil aprender o que os matemáticos fazem, os vários tipos de cálculos, do que tentar buscar uma definição, que pode ser vazia, se não consideramos a grande diversidade de tópicos tratados na disciplina.

1. O que é a metafísica?

A metafísica é uma das disciplinas da filosofia, como a epistemologia, a ética, a estética, entre muitas outras disciplinas que poderíamos destacar aqui. Podemos dizer que a metafísica é a mais geral e, talvez, mais central das

disciplinas da filosofia. Ela tem uma longa história que remonta aos filósofos pré-socráticos e à tentativa de determinar quais são as entidades mais fundamentais que existem, quais são os elementos que compõem a realidade, ou ainda, qual o princípio ordenador das coisas. Encontramos também um pensamento metafísico em filósofos como Platão. A sua famosa teoria das ideias (ou formas) é sem dúvida uma teoria profundamente metafísica, e parece que é a primeira tentativa de resolver o que chamamos aqui de “problema dos universais”. Mas o primeiro filósofo a tentar definir o que é a metafísica e quais os objetivos dessa disciplina foi Aristóteles, na obra intitulada *Metafísica*. A obra foi assim nomeada posteriormente, por Andrônico de Rodes, que foi quem organizou as obras de Aristóteles por volta do séc. I a.c. O título se refere mais ao modo como estão organizadas as obras – depois da física – do que a algo relacionado ao conteúdo da obra.

Aristóteles chamou a disciplina que se encontra na *Metafísica* de “filosofia primeira” ou “teologia”. Ele afirma que o objetivo central dessa disciplina é alcançar a *sabedoria*. Contudo, ao tentar caracterizar a metafísica, Aristóteles não nos dá uma visão unívoca do que seria tal disciplina, mas nos fornece duas caracterizações. Em uma das caracterizações, afirma que o que se busca é o conhecimento das causas primeiras (*Meta.*, A.1). Neste sentido, o objetivo que teríamos, ao fazer metafísica, seria o conhecimento de Deus ou do Motor Imóvel. O tema central da metafísica nessa perspectiva é, portanto, a “substância imaterial” (*Meta.*, E.1).

Contudo Aristóteles não caracteriza a metafísica apenas como a ciência das causas primeiras, mas também é a ciência do *ser enquanto ser* (*Meta.* Γ.1). Nessa segunda caracterização, a metafísica é entendida como uma ciência em busca de investigar o ser enquanto algo que simplesmente existe. Assim, busca-se estabelecer quais as características ou atributos que essas entidades têm de exibir, simplesmente enquanto seres existentes. Assim, a metafísica será vista como uma teoria sobre categorias, que busca estabelecer quais são as categorias mais gerais e objetivamente existentes, nas quais caem todos os seres existentes – e talvez também os seres possivelmente existentes. Essas categorias são também conhecidas por *categorias ontológicas*. Uma discussão muito influente sobre quais são as categorias ontológicas está no tratado de

Aristóteles intitulado *Categorias*⁷. Algumas metáforas são comumente usadas para falar dessas disputas sobre categorias ontológicas; por exemplo, uma delas diz que ao fazermos uma teoria das categorias queremos “cortar a realidade nas juntas”, outra é a de que estamos em busca do “alfabeto do ser”.

Mas será que a metafísica é uma disciplina que tem esses dois objetivos diferentes? Será preciso ter uma concepção não unívoca da disciplina? Há uma tensão aqui entre essas duas concepções? Aristóteles pensa que essa tensão entre as duas concepções é meramente aparente. Ele afirma que a ciência das causas primeiras identificará as causas subjacentes às características primárias das coisas. Assim, a ciência que estuda as causas primeiras é simplesmente a ciência que estuda o ser enquanto ser. Além disso, Aristóteles parece defender que, uma vez que a disciplina examina todas as coisas enquanto seres existentes, estará Deus ou o Motor Imóvel entre as coisas que merecem ser caracterizadas⁸ (*Meta*. E.1).

Nesse contexto da metafísica de Aristóteles, surge um conceito muito importante para a disciplina: essência. O conceito é importante, juntamente com outros conceitos metafísicos, na medida em que podemos pensar que a tarefa da metafísica é a de descobrir a essência dos seres ou das coisas, o que eles são em si mesmos, independentemente de nosso pensamento e linguagem. Essa perspectiva aristotélica da metafísica foi em grande parte abraçada pelos filósofos medievais. Eles buscaram defender que a metafísica era uma disciplina preocupada a natureza ou essência da substância divina e, ao mesmo tempo, uma disciplina cujo objetivo é estabelecer as categorias ontológicas mais fundamentais.

Os filósofos racionalistas modernos – a partir do séc. XVI – buscaram ampliar o âmbito da metafísica. Muitos assumiram a distinção entre a *metafísica geral* e a *metafísica especial*. A metafísica geral, posteriormente conhecida por *ontologia*, tem por objetivo estabelecer quais são as *categorias ontológicas*, no sentido aristotélico do termo. Entretanto, a metafísica especial vai tratar de três

⁷O tratado, juntamente com os tratados *Da Interpretação*, *Analíticos Anteriores*, *Analíticos Posteriores*, *Tópicos* e *Refutações Sofísticas*, compõem o chamado *Órganon*, que é o conjunto de obras de Aristóteles sobre lógica.

⁸Para uma discussão um pouco mais aprofundada do tópico sobre a caracterização de Aristóteles da metafísica ver Loux (2006).

seres específicos e certos conceitos relacionados a esses seres, gerando assim três subdisciplinas: a cosmologia, que trata do universo e da causalidade que há nele; a psicologia racional, que trata da alma humana (ou mente, como se prefere hoje em dia) e a sua relação com o corpo; e a teologia natural, que trata da existência e natureza de Deus.

Os racionalistas modernos levaram teorização metafísica às últimas consequências. Exemplos disso vemos em filósofos como Leibniz e Spinoza, que desenvolveram sistemas metafísicos altamente abstratos e que parecem muito longe do senso comum. Isso levou alguns filósofos a serem críticos ferrenhos da teorização metafísica. Entre eles podemos destacar Hume e Kant. Hume é um representante do empirismo britânico, que é a escola filosófica que rivaliza com o racionalismo. A oposição entre empirismo e racionalismo é em grande parte uma distinção relacionada à epistemologia⁹ e ao modo como conhecemos o mundo. Mas essa distinção tem consequências para a metafísica, e muitas vezes motiva posições antimetafísicas. Tanto a crítica de Hume quanto a crítica de Kant à metafísica parecem em parte ter uma motivação epistemológica.

No livro *Investigação acerca do entendimento humano* (1748), Hume promove críticas ao modo *a priori* como se desenvolve a metafísica. Ele é um empirista, de modo que pensa que todo conhecimento genuíno deriva da experiência sensível, ou seja, é *a posteriori*. Entretanto, os racionalistas pensavam que podemos ter conhecimento genuíno e *a priori*, um conhecimento que é independente da experiência sensível, adquirido apenas por raciocínio. O exemplo mais óbvio é o que se tem em matemática. E talvez seja esse o modelo de conhecimento almejado pelos racionalistas. Mas mesmo em relação à noção de causa e efeito, parecem pensar os racionalistas, poderíamos deduzir *a priori* uma conexão necessária entre um evento que é causa e outro evento que é efeito. Mas Hume¹⁰ procurou argumentar que, uma vez que não temos a experiência de conexões necessárias, mas apenas vemos eventos se seguirem de outros eventos em uma conjunção constante, é equivocado pressupor que

⁹A epistemologia é uma disciplina da filosofia que trata do nosso conhecimento da realidade. Juntamente com a metafísica, ela é uma das mais gerais disciplinas da filosofia.

¹⁰Principalmente, seções 4 e 7 do livro *Investigação acerca do entendimento humano*.

haja alguma conexão necessária entre eventos que são causas e eventos que são efeitos. Hume é sem dúvida um dos filósofos mais influentes entre os filósofos contemporâneos. Tanto que ainda hoje temos teorias humeanas sobre a causalidade, sobre as leis da natureza e sobre tópicos relacionados. As teorias humeanas podem ser vistas como uma posição metafísica, mas uma posição metafísica deflacionária, na medida em que evita compromisso com certos tipos de entidades; em particular, com entidades ditas “intensionais”¹¹.

Kant vem após Hume e sob a sua influência. Kant foi o primeiro filósofo a fazer a interessante pergunta: a metafísica é possível? Em sua principal obra, *a Crítica da razão pura* (1781, 1787), Kant tem por objetivo mostrar em que medida a metafísica é possível¹². Ele se afasta completamente da tradição aristotélica e racionalista: o conhecimento da essência do ser, da coisa-em-si, não é possível. Podemos conhecer apenas os fenômenos nos quais figuram tais coisas, não as coisas em si mesmas. A tarefa da metafísica, de acordo com Kant, não é a de estabelecer as categorias mais gerais da realidade, mas delinear os conceitos inatos da faculdade cognitiva humana, a partir dos quais apreendemos os fenômenos. A ideia básica é temos certos dados dos sentidos que são de alguma forma organizados por esses conceitos inatos. Assim, quando estamos em contato com uma coisa, não apreendemos a coisa mesma, mas uma representação dessa coisa; representação essa que é mediada por aqueles conceitos inatos. A metafísica busca estabelecer aquilo que Kant chamou de “verdades sintéticas *a priori*”.

¹¹ Muitos filósofos da tradição humeana têm aversão à postulação de entidades intensionais em suas ontologias. A distinção entre entidades extensionais e intensionais pode ser colocada a partir da noção de conjunto: um conjunto pode ser determinado pelos seus membros (extensões), ou por meio de uma propriedade possuída pelos seus membros (intensão). Por exemplo, o conjunto de sábios (S). Podemos determiná-lo extensionalmente, especificando cada um dos seus membros S: {Sócrates, Platão, Aristóteles...}; ou podemos determiná-lo intencionalmente, especificando a propriedade que é possuída por cada um dos seus membros S: { $x \in S$ | x é sábio}. No primeiro caso, para especificar o conjunto, recorreremos a entidades extensionais, os indivíduos Sócrates, Platão, Aristóteles (...); no segundo caso, recorreremos a uma entidade intensional, que é a propriedade de ser sábio.

¹²Ver o prefácio à segunda edição da *Crítica da Razão Pura* (1787) de Kant, onde ele expõe esse ponto.

Uma pergunta óbvia que poderia ser feita aqui por um aristotélico ou um racionalista é: mas se não podemos conhecer as coisas-em-si, como podemos conhecer os processos internos e essenciais à nossa faculdade cognitiva? Nossa capacidade cognitiva é apenas mais um item do mundo, e que não parece que estamos mais aptos a conhecer a sua essência, ou o seu em-si, do que a essência de uma determinada árvore ou átomo, ou qualquer outra entidade no universo. Assim, podemos não estar dispostos a aceitar a delimitação proposta por Kant para os limites da metafísica. Eu, particularmente, não estou.

No século XX, de modo mais forte na primeira metade do século, temos uma grande tradição antimetafísica. “Metafísico” passou a ser uma espécie de xingamento. Essa tradição que já vem de Hume, se intensificou devido à grande influência de investigações relacionadas à linguagem, que passaram a ser mais centrais na filosofia, seja na tradição analítica, seja na tradição continental¹³. Diz-se comumente que houve uma *virada linguística* na filosofia nesse momento histórico. Instigações sobre a natureza das coisas passaram a investigações sobre a linguagem e ao modo como falar sobre as coisas. Podemos destacar aqui pelo menos três posições antimetafísicas que se desenvolveram nesse contexto: o relativismo, o cientificismo e o neokantismo.

O relativismo é a defesa do que podemos chamar de “relativismo global”¹⁴. A ideia básica é a de a noção de verdade objetiva não faz sentido, toda verdade seria relativa a algo – ao modo como *escolhemos* categorizar o mundo,

¹³ Não há uma definição especificando a diferença entre a filosofia analítica e a continental. Acredito que a distinção deve, grosso modo, ser entendida em termos temporais, geográficos e metodológicos. A distinção é temporal na medida em que engloba basicamente filósofos do século XX. A distinção é geográfica na medida em que a filosofia analítica é a filosofia feita inicialmente nos países de língua inglesa, enquanto a filosofia continental era feita na Europa continental, principalmente na França e Alemanha. Mas não se pode generalizar, pois uma importante escola da filosofia analítica floresceu na Alemanha na e Áustria: o positivismo lógico. Mas esses estavam interessados nos problemas e métodos desenvolvidos pelos filósofos analíticos insipientes. Nesse sentido é que a distinção também se dá em termos metodológicos. Esta introdução segue uma linha analítica, tratando da metafísica dentro dessa tradição. Sobre a história da filosofia analítica, uma boa introdução é Stephen P. Schwartz (2017).

¹⁴O filósofo francês Jaques Derrida, defensor do desconstrucionismo, é considerado um filósofo relativista.

por exemplo. Podemos falar de “relativismo local”, que seria o relativismo restrito a uma área específica, à ética, por exemplo, mas esse não é relevante para colocar objeção à metafísica. Aqui será relevante o relativismo que recusa a própria noção de verdade objetiva. Isso porque, tanto a metafísica aristotélica quanto a metafísica racionalista pressupõem uma noção de verdade objetiva, não relativa. Acredito que o relativismo é profundamente incoerente, ou o suficiente para não recusarmos a metafísica baseados nessa posição. A tese do relativismo global pode ser expressa pela frase “Não há verdades absolutas”. Podemos nos perguntar: aquilo que essa frase expressa é uma verdade absoluta? Se é, há verdades absolutas (há pelo menos essa!). Se não é, segue-se que o relativismo é incorreto e, que é no mínimo possível haver verdades que sejam absolutamente verdadeiras. Em ambos os casos as coisas estão más para o relativista.

Sobre o cientificismo, gostaria também de falar algumas palavras¹⁵. A tese do cientificismo pode ser resumida do seguinte modo: “a única metafísica legítima é aquela que está em continuidade com a ciência empírica”. Podemos dizer também que defendem que o que há é aquilo que a nossa melhor ciência diz que há. Buscando ser breve aqui também, podemos perguntar: mas essa tese está em continuidade com a ciência empírica? Parece que não: não há testes empíricos ou observações possíveis que sugiram que essa tese seja empiricamente adequada. Ela simplesmente não é empírica, está além da ciência empírica. E nesse sentido deveria ser recusada pelos próprios critérios. A conclusão aqui é a de que parece possível fazer metafísica de uma forma genuína sem estar em continuidade com a ciência empírica – ou melhor, há domínios da metafísica que não dizem respeito à ciência empírica. Um exemplo óbvio é a discussão sobre entidades abstratas¹⁶. Saber se existem ou não tais entidades é uma questão completamente não empírica.

Ainda uma outra posição antimetafísica é o neokantismo. A ideia básica

¹⁵Os positivistas lógicos do Ciclo de Viena podem ser vistos como defensores dessa posição.

¹⁶Entidades abstratas se opõem a entidades concretas, e uma discussão importante é entender que tipo de coisa são essas entidades, tanto as concretas quanto as abstratas. A distinção é por vezes entendida no sentido espaciotemporal: entidades concretas são aquelas que estão no espaço e no tempo, enquanto as abstratas são aquelas sem localização espaciotemporal.

desses filósofos é de que a metafísica não tem por objetivo caracterizar as categorias mais gerais da realidade, mas de delinear nossos esquemas conceituais por meio dos quais compreendemos o mundo. A diferença aqui com o próprio Kant, é que para os filósofos nessa vertente pode haver vários esquemas conceituais, e não apenas um, como parece supor Kant. Um bom exemplo desse tipo de investigação é Strawson (1959). Ele busca fazer o que ele chama de “metafísica descritiva”, cujo objetivo é justamente delinear aqueles conceitos mais gerais. Alguns críticos pensam que Strawson vem com essa conversa sobre a metafísica descritiva, mas o que ele de fato faz é uma metafísica aristotélica pura e simples. Era um metafísico escondido no armário.

A partir dos anos 70 do século XX houve o que podemos chamar de uma “virada ontológica”. Por influência de filósofos como Saul Kripke e Hilary Putnam. Kripke (1980) argumentou a favor da existência de verdades necessárias conhecidas *a posteriori*. Desde a filosofia moderna tanto racionalistas quanto empiristas tinham uma tendência a identificar, por um lado, aquilo que é uma verdade necessária com as verdades que são conhecíveis *a priori* e com as verdades analíticas; e por outro lado, identificavam verdades contingentes, com verdades *a posteriori* e verdades sintéticas. Kant foi o primeiro filósofo a questionar essa identificação, propondo que existiam verdades sintéticas *a priori*. Mas o que Kripke fez foi bastante revolucionário: buscou mostrar que as modalidades necessário/contingente, *a priori/a posteriori* e analítico/sintético, são modalidades diferentes, metafísicas, epistêmicas e semânticas, respectivamente. Dessa forma, procurou mostrar que há verdades que são necessárias e são conhecidas *a posteriori*, e que também há verdades que são contingentes que são conhecidas *a priori*.

A ideia de que há verdade necessárias *a posteriori* é de grande importância para a ciência empírica: mostra que nem todo conhecimento empírico, baseado na experiência, é contingente. Um exemplo de verdades científicas que parecem necessárias são certas identificações teóricas como “a água é H₂O”, ou “Sócrates é da espécie humana”. Se essas frases são de fato verdadeiras, segue-se, de acordo com Kripke, que são necessariamente verdadeiras. Isso quer dizer que não há mundos (metafisicamente) possíveis¹⁷

¹⁷A noção de mundo possível foi proposta por Leibniz. Com ela queremos falar como o mundo é, e como poderia ser. Hoje é muito utilizada na lógica modal e na metafísica

nos quais temos água e não temos H₂O, por exemplo. Os mundos nos quais temos água, mas não temos H₂O seriam mundos meramente logicamente possíveis, e não mundos realmente possíveis. E parece que nem tudo que é logicamente possível é de fato possível: é logicamente possível, por exemplo, Sócrates ser um chinelo, mas ninguém que esteja fora de um hospício por mérito diria que Sócrates, ao invés de ser uma pessoa, poderia ter de sido de fato um chinelo. Pensamos que Sócrates poderia não ter sido um filósofo, mas parece obviamente errado pensar que ele poderia não ser da espécie humana.

A ideia de que há verdades contingentes que são conhecidas *a priori* não é relevante para a ciência, mas nos ajuda a mostrar que não podemos identificar, por um lado, as noções de verdade necessária e verdade *a priori* e, por outro lado, as noções de verdade contingente e verdade *a posteriori*. Um exemplo de uma verdade contingente e *a priori* seria aquela expressa no cogito cartesiano: “Eu existo”. Essa é uma verdade *a priori*, na medida em que a descubro apenas pelo raciocínio (quando digo a frase não posso estar errado!); mas a frase expressa uma verdade contingente, na medida em que eu não existo necessariamente – existo mas poderia não ter existido, o que teria sido uma coisa bem triste.

Com essa *virada ontológica*, os metafísicos puderam sair do armário. Deixou de ser visto como um absurdo fazer metafísica e ela voltou a ser uma disciplina respeitada nos departamentos de filosofia de influência analítica, principalmente. Isso não foi o fim das posições antimetafísicas, mas sim o florescimento de muitas teorias de influência aristotélica, da metafísica como a ciência do ser enquanto ser, que gozava de pouca popularidade até o momento.

Hoje, devido às especificidades que foram tomando as disciplinas da filosofia, a metafísica já não é uma disciplina tão ampla como era para os racionalistas modernos. A questão sobre a existência de Deus, por exemplo, é debatida na filosofia da religião, que reúne questões metafísicas, epistemológicas e morais relacionadas à religião. A relação entre o corpo e a

como uma forma de expressar as noções de necessidade, possibilidade e contingência. Por exemplo, a proposição expressa por “ $2+2=4$ ” é necessariamente verdadeira, uma vez que é verdadeira em todos os mundos possíveis. A proposição expressa por “Sócrates é filósofo” é contingentemente verdadeira, uma vez que é verdadeira no mundo atual, mas é falsa em algum mundo possível.

mente é estudada pela filosofia da mente, que também reúne um conjunto de problemas relacionados à natureza da mente, não apenas problemas metafísicos. A cosmologia, por sua vez, é mais uma área da física, ou da astrofísica, do que propriamente da filosofia; entretanto, ela não se abdicou de seu caráter altamente metafísico na busca de compreender origem e essência do universo.

A metafísica atualmente é entendida mais no sentido aristotélico da ciência do ser enquanto ser, ou como a metafísica geral dos racionalistas. De modo que tratar de questões sobre categorias ontológicas e a aplicabilidade de tais categorias para resolver problemas relacionados a outras entidades é a principal tarefa dos metafísicos contemporâneos. Assim, temos problemas como o problema dos universais e o problema dos particulares, nos quais buscamos determinar, entre outras coisas, quais as entidades mais básicas que compõem a realidade, o alfabeto do ser. Problemas sobre a persistência das entidades particulares ao longo do tempo. Problemas relacionados a certas noções como, tempo, proposições, eventos, causalidade, leis da natureza, possibilidade, necessidade, mundos possíveis, entre outros conceitos importantes.

Uma área dentro da metafísica que tem atraído muitos filósofos atualmente é o que podemos chamar de “metametáfísica”. Nessa área, estão em discussão conceitos fundamentais da metafísica. Por exemplo, queremos entender o que são categorias ontológicas, ou o que são certas relações ontológicas como a de fundação (do inglês *grounding*), dependência ontológica, entre outros. São questionamentos, não sobre a natureza da realidade, mas sobre os conceitos fundamentais que usamos para falar dessa realidade. Na metametáfísica busca-se fundamentar a própria metafísica.

Como nosso objetivo aqui é introdutório, procurarei falar aqui sobre quatro problemas que considero centrais na discussão metafísica e que podem ser úteis para alguém que deseja fazer uma incursão a essa disciplina. Procurarei indicar alguma bibliografia relevante para a estudos posteriores. Os problemas que vos apresento brevemente são: o problema dos universais, o problema sobre os particulares concretos, o problema sobre a natureza do tempo e, por fim, o problema sobre a persistência dos particulares concretos ao longo do tempo.

2. O problema dos universais

Um dos problemas metafísicos mais antigos e mais debatidos ao longo da história da filosofia é o problema dos universais. Definir exatamente qual é o problema dos universais não é simples e já é em si um problema filosófico¹⁸. Além disso, quando falamos em “problema dos universais”, podemos ser levados a pensar que seja algum problema relacionado aos universais – e há muitos –, mas como veremos, falar em universais já é uma tentativa de se resolver o problema que está em questão.

Uma caracterização mínima e talvez aceitável sobre esse problema seria dizer que ele versa especialmente sobre a categoria dos atributos, que é composta pelas propriedades, relações e, para algumas teorias, por espécies naturais¹⁹. Por exemplo, quando dizemos que Sócrates é sábio, estamos predicando a sabedoria ao indivíduo Sócrates. A sabedoria é uma propriedade²⁰. Quando dizemos que Sócrates é mestre de Platão, estamos predicando *ser mestre de* ao par de indivíduos Sócrates e Platão. *Ser mestre de* é uma relação. E quando dizemos que Sócrates é humano, estamos predicando a humanidade a Sócrates. E a humanidade (ou a espécie humana) é uma espécie natural.

Há várias perguntas interessantes a se fazer sobre os atributos. Podemos nos perguntar sobre a existência e natureza dos atributos. São eles entidades *sui generis*, i.e., entidades irreduzíveis a outros tipos de entidades mais básicas? Se são entidades *sui generis*, que tipo de entidades seriam essas? Se não são entidades *sui generis*, como compreender o discurso sobre atributos?

¹⁸Para uma discussão interessante e pouco convencional do problema, ver Rodriguez-Pereyra (2000).

¹⁹As espécies naturais são universais expressos por “humano”, “ouro” etc. Esse tipo de universal, segundo alguns filósofos, é irreduzível a propriedades e relações. Essa é a posição de Aristóteles nas *Categorias* e de Lowe (2006). Aristóteles chama às entidades dessa categoria de “substâncias secundárias”.

²⁰Nem toda teoria que admite universais vai admitir que a sabedoria é uma propriedade genuína. Usei e usarei o exemplo por simplicidade. Entretanto, tenha em mente que alguns filósofos, como Armstrong (1978b, 1989), defendem que a última palavra sobre quais universais realmente existem é da ciência, especialmente a física. Para essa teoria dos universais, o universal da sabedoria não existe, mas existe, por exemplo, o universal de ter carga -1, instanciado por todos os elétrons.

Nesse último caso, vamos precisar de alguma noção que substitua adequadamente os atributos. Além disso, há uma relação desse problema sobre os atributos com outros tópicos filosóficos e que merecem ser considerados: a predicação, a referência a termos abstratos singulares, as proposições, a natureza dos particulares concretos, os eventos, a causalidade, as leis da natureza, a natureza dos mundos possíveis, entre outros. Na verdade, a depender da resposta que se dá ao problema dos universais, haverá muitas consequências para as teorizações posteriores nesses diversos tópicos. Talvez por isso seja interessante começarmos com o problema dos universais, apesar de não haver uma obrigação de assim fazer.

Mas como surge o problema dos universais? Acredito que o problema surge da necessidade de se explicar certos fatos sobre o mundo: o fato do indivíduo Sócrates ser sábio, o fato do indivíduo Platão ser sábio, o fato de Sócrates ser mestre de Platão, o fato de Platão ser mestre de Aristóteles, o fato dos elétrons terem todos eles carga negativa (-1), entre uma infinidade de fatos que poderíamos incluir aqui nessa lista, mas que não seria conveniente. O ponto central é que temos indivíduos (ou quase-indivíduos, no caso dos elétrons e outras partículas subatômicas) exibindo certos atributos; queremos saber afinal de contas o que são esses atributos e qual a natureza deles. Falar algo sobre a natureza dessas supostas entidades nos ajuda a entender em parte os fatos que levantaram o problema. Além disso, falar algo sobre a natureza e utilidade dessas entidades envolve um vasto conjunto de problemas metafísicos, lógicos e epistemológicos. De modo que, novamente, falar do problema dos universais pode ser errôneo, pois parece haver um problema específico, quando na verdade envolve muitos problemas relacionados entre si.

Sem dúvida, um dos problemas mais relevantes que se levanta é o problema do *um sobre muitos*. O problema do *um sobre muitos* gira em torno da pergunta sobre como são possíveis conjuntamente fatos como Sócrates é sábio, Platão é sábio, Aristóteles é sábio etc. Não estamos questionando a existência desses indivíduos (Sócrates, Platão e Aristóteles), mas queremos saber como eles podem ao mesmo tempo ter essa mesma propriedade de ser sábio – tomando o termo “propriedade” em um sentido pré-teórico, sem avançar em alguma teoria sobre propriedades. O que é essa propriedade? É a mesma propriedade que ocorre nos três fatos? Serão diversas propriedades particulares, uma para cada caso? Será que temos que falar de propriedades em um sentido

substancial do termo, ou podemos ter uma noção menos substancial, ou mesmo deflacionista sobre as propriedades? As mesmas perguntas valendo para os demais atributos.

Os filósofos realistas defendem que para solucionarmos o problema do *um sobre muitos* e outros problemas relacionados, precisamos de uma concepção substancial de atributos. Temos que entender essas entidades como propriedades e relações *sui generis*, irreduzíveis a outros tipos de entidades mais básicas. Em outras palavras, temos que entender essas entidades como compondo uma categoria ontológica fundamental. Os realistas pensam ainda que temos que entender essas entidades como sendo universais. Na literatura sobre o tema, parece que uma condição suficiente para uma entidade ser universal é essa entidade poder ter múltiplas instâncias ao mesmo tempo. Nesse sentido, a propriedade da sabedoria será uma propriedade universal que tem como instâncias Sócrates, Platão, Aristóteles, e os demais sábios.

Historicamente, o realismo dos universais tem a sua origem nos filósofos Platão²¹ e Aristóteles²². As concepções desses filósofos podem ser chamadas de “realistas”, mas são essencialmente diferentes. Essa diferença é notada ainda hoje entre as teorias platônicas (ou platonistas, como preferirem) e aristotélicas. Platônicos e aristotélicos divergem sobre quais universais realmente existem²³, sobre a natureza dos universais, sobre como conhecemos tais entidades, entre outras divergências mais pontuais. É preciso ter cuidado ao falar de teorias platônicas e aristotélicas, na medida em que não estamos falando especificamente das teorias defendidas por Platão e Aristóteles, mas de teorias que foram defendidas e que em traços gerais e relevantes se assemelham às teorias daqueles filósofos.

Podemos destacar aqui algumas divergências que há entre as teorias

²¹Para saber mais sobre o modo como Platão trata os universais (ou as formas), ver principalmente o diálogo *Parmênides*, onde ele expõe a teoria das formas e são levantadas várias objeções a essa teoria, como o famoso *argumento do terceiro homem*. Ele se refere a essas entidades em outros diálogos (e.g., *A República*), mas apenas no *Parmênides* elas são o foco da discussão.

²²Obras importantes de Aristóteles nas quais eles trata do tema são *As Categorias* e *AMetáfisica*.

²³Há tipos diferentes de teorias platônicas e aristotélicas, que vão divergir internamente sobre essas questões.

platônicas e aristotélicas. Para os platônicos os universais são entidades não espaciotemporais, abstratas nesse sentido. Para os aristotélicos os universais têm algum tipo de relação espaciotemporal: estão localizados em muitos lugares ao mesmo tempo. São abstratas em um sentido mais fraco do termo. Os universais platônicos existem independentemente de haver instâncias suas: o universal da sabedoria existe mesmo que não exista coisas sábias. Enquanto os universais aristotélicos existem de forma dependente de suas instâncias: o universal da sabedoria existe apenas na medida em que há coisas sábias. Na teoria aristotélica podemos dizer que há uma prioridade dos objetos concretos (particulares) em relação aos abstratos (universais), que existem em uma relação de dependência ontológica em relação aos objetos concretos.

Na metafísica contemporânea o platonismo é representado por filósofos como Russell (1912), Loux (1978) e Tooley (1977, 1987). Os dois primeiros em uma vertente mais semanticista, preocupados com a predicação e a referência abstrata. Tooley está mais envolvido em um debate sobre a causalidade e sobre o que são leis da natureza²⁴. Sobre as teorias aristotélicas, temos uma maior variedade de posições possíveis. Mas para citar dois nomes muito relevantes para o debate, temos Armstrong (1978a, 1978b, 1989) e Lowe (2006).

Entretanto há filósofos que não estão satisfeitos com o realismo dos universais. São os chamados “nominalistas”. O nominalismo é ainda mais heterogêneo, reunindo sob essa alcunha teorias que não têm muito em comum. Um acordo comum entre esses filósofos é que o que existe são entidades particulares, mas não entidades universais. Contudo, eles não estão de acordo sobre quais entidades particulares realmente existem.

Historicamente, um grande nome do nominalismo é William de Ockham (Ver Spade, 1994). Este defende, em sua ontologia, a existência de substâncias, que são certo tipo de indivíduos ou particulares concretos, e qualidades, que são aquelas entidades que estamos chamando de “propriedades”. Mas esse tipo de nominalismo pressupõe que as propriedades são particulares, no sentido em que estão intimamente ligadas à sua instância. Retomando o nosso exemplo, há a propriedade da sabedoria de Sócrates e a propriedade da sabedoria de Platão, mas essas propriedades são cada uma delas uma ocorrência particular, e não a mesma propriedade que ocorre em Sócrates e Platão, como supõe o realismo.

²⁴Sobre o tópico universais e leis da natureza, ver Cid (2019).

Esse tipo de teoria é hoje conhecido como *teoria dos tropos*. Há diversas variantes, mas todas concordam que há atributos como entidades *sui generis*, entretanto esses devem ser entendidos como particulares, não como universais. Quem denominou os atributos particulares como “tropos” foi Williams (1953), em uma brilhante defesa dessa teoria. Outros defensores são Stout (1923) e Campbell (1990). Este último chama a essas entidades de “particulares abstratos”. Outro termo por vezes utilizado para falar dos tropos é “acidentes”.

Outros nominalistas buscam uma estratégia deflacionária em relação aos atributos: essas entidades não existem, ou pelo menos não precisamos nos comprometer com a existência delas. O nominalismo da semelhança defende que há indivíduos e relações de semelhança entre esses indivíduos. Nesse sentido, dizemos que Sócrates e Platão existem e se assemelham em alguma medida, e por isso podemos dizer que são sábios. Por medida de espaço aqui não entrarei em mais detalhes sobre a teoria. Ela foi defendida pelo filósofo Gonzalo Rodriguez-Pereyra (2002).

Na mesma direção vai a teoria nominalista das classes. Esse tipo de teoria diz que a noção de atributos se reduz à noção de classes. Nos fatos de que Sócrates é sábio e de que Platão é sábio, temos os indivíduos Sócrates e Platão e esses são membros da classe de pessoas sábias, ou dos sábios. Por isso podemos predicar “sábio” àqueles indivíduos. Essa teoria é levada às últimas consequências por David Lewis (1983, 1986). Lewis pensa que há uma pluralidade de mundos possíveis concretos, e essas classes estão espalhadas por esses mundos. De modo que a classe de sábios inclui os sábios de todos os mundos possíveis, e não apenas do mundo atual – esse contorcionismo todo para evitar problemas com atributos co-instanciados²⁵.

Uma acusação muito comum que se faz aos filósofos realistas é que suas teorias não explicam os fenômenos que se propõem, postulando entidades que por si mesmas já são duvidosas, ou faltam-lhe o poder explicativo requerido pelos realistas. Por outro lado, o nominalismo, principalmente os dois últimos citados, invertem a ordem da explicação: não é porque Sócrates e Platão se

²⁵Isto é, atributos que parecem diferentes, mas são instanciados pelos indivíduos objetos no mundo atual. Se falamos apenas de classes do mundo atual, dada a extensionalidade dos conjuntos, teremos que tomar esses atributos como sendo os mesmos atributos, ainda que isso contrarie a nossa intuição de que são atributos diferentes.

assemelham ou pertencem ao conjunto dos sábios, que eles são sábios; mas antes, o fato de eles serem sábios é que explica o fato deles se assemelharem ou de pertencerem ao conjunto dos sábios.

Mas cada lado da discussão tem também intuições que o favorecem. De um lado, o realismo pode satisfazer o nosso desejo por uma análise mais profunda, mas recorrendo a entidades “duvidosas” e que talvez expliquem as coisas apenas aparentemente. De outro lado, o nominalismo parece não explicar muito bem as coisas, ou deixar muitas delas intencionalmente sem explicação, mas tem uma ontologia²⁶ simples e com entidades intuitivamente mais aceitáveis do que os realistas. A escolha aqui é difícil, e dependerá da ideologia de cada um e do quanto estará disposto a pagar, ao assumir uma ontologia. Afinal, não existe almoço grátis.

3. O problema sobre os particulares concretos

A existência dos indivíduos, também chamados “particulares concretos”, a princípio não parece levantar problemas ontológicos muito graves. Sabemos, por exemplo, que o indivíduo chamado “Sócrates” existe (ou existiu, para ser mais preciso), e que Platão também existe, e assim por diante. Dessa forma, não temos um problema sobre a existência dessas entidades. O problema surge quando vamos nos perguntar sobre a natureza delas. Em que medida esses indivíduos são portadores de atributos? Como esses indivíduos se relacionam com os seus atributos? Os indivíduos são entidades constituídas por seus atributos e talvez alguma outra entidade? Ou devem ser entendidos como entidades básicas, como formando uma categoria ontológica *sui generis*?

A noção que inicialmente parecia não levantar problemas, agora nos coloca diante de um grande problema, que é o de dar um tratamento adequado a essas entidades. Obviamente, aquilo que dissermos sobre o problema dos

²⁶Há uma distinção entre ontologia e ideologia que está por trás de toda teoria ontológica: por um lado temos uma ontologia, que são as entidades que a teoria diz que existe, com as quais está comprometida. Por outro lado, temos uma ideologia, que são certos compromissos com determinadas ideias que essas teorias também têm; por exemplo, podemos ser materialistas, o que terá grande influência sobre nossa teoria ontológica e mesmo com o tipo de entidade que estamos dispostos a admitir como existentes nessas teorias.

universais terá reflexos aqui nessa discussão, uma vez que um dos pontos aqui é saber qual é a relação entre os indivíduos e os atributos que eles exibem.

A resposta dos filósofos nominalistas das classes e da semelhança parece fácil: dirão simplesmente que os indivíduos são entidades brutas, não passíveis de uma análise ontológica em termos de entidades mais básicas. Dessa forma, o debate sobre a natureza dos particulares concretos termina prematuramente, se estamos dispostos a aceitar alguma teoria nominalista que não admite a existência de atributos como entidades *sui generis*. Mas como vimos brevemente, há um tipo de teoria nominalista que admite a existência de atributos: a teoria dos tropos.

Os filósofos que postulam atributos como entidades *sui generis* precisam dizer de que maneira esses atributos se relacionam com os seus portadores. Uma teoria platônica também tende a não analisar os particulares em termos de suas propriedades. Acredito que há duas opções aqui: tomar os particulares concretos como primitivos e mostrar que há algum tipo de relação ontológica desses particulares com os universais abstratos – relação de participação, ou de exemplificação –; ou analisar esses particulares em termos de feixes de tropos – o que veremos mais adiante –, e dizer que esses tropos têm alguma relação ontológica com os universais platônicos.

Um realista aristotélico tem opções de teorias sobre particulares concretos. Em uma das opções, eles podem entender tais entidades, ou pelo menos uma parte considerável delas, como substâncias aristotélicas, que são entidades *sui generis*, irreduzíveis a outros tipos de entidades, e das quais todas as demais entidades que existem dependem para as suas existências. Os atributos dessas substâncias constituem outras categorias ontológicas também fundamentais, mas todos eles dependentes da substância para as suas existências. Algumas teorias vão distinguir pelo menos quatro tipos de entidades fundamentais e irreduzíveis: substâncias, atributos (propriedades e relações universais), espécies naturais e acidentes (tropos). Uma teoria com quatro categorias fundamentais é defendida por Lowe (2006). É argumentável que Aristóteles nas *Categorias* tenha uma teoria das quatro categorias.

A outra opção para o aristotélico é analisar os indivíduos em termos de substratos e propriedades intrínsecas²⁷, tomando essas entidades como básicas.

²⁷Uma propriedade intrínseca é aquela que um indivíduo tem em virtude de sua natureza

Consideremos o indivíduo Sócrates. Ele seria uma entidade redutível, pelo menos em tese, a um substrato mais suas propriedades. O substrato seria uma espécie de sustentáculo dos atributos, o que está debaixo dos atributos, sendo que ele próprio não tem qualquer atributo. Por isso são por vezes conhecidos como “particulares nus”, os particulares despidos de todas as suas propriedades. As propriedades aqui podem ser entendidas como universais (aristotélicos) ou como tropos. Se entendemos as propriedades como universais, uma das funções do substrato é conferir a particularidade aos indivíduos – se fossem constituídos apenas de universais, seria metafisicamente possível haver dois indivíduos indiscerníveis²⁸. Esse é o problema sobre a particularidade dos indivíduos. A teoria que postula substratos resolve facilmente esse problema, a troco de postular uma entidade que não sabemos exatamente o que é – quando conhecemos algo, conhecemos as suas propriedades e relações, o que os substratos, por definição, não têm. Uma teoria mais ou menos nesses moldes foi defendida por Armstrong (1978b, 1989), tomando as propriedades como universais aristotélicos. Tomando as propriedades como tropos, o que é menos comum, temos filósofos como Martin (1980) e Heil (2003).

Outra estratégia é a da *teoria dos feixes*. Essa teoria é a defesa de que um particular concreto como Sócrates não seria algo além de um feixe de suas propriedades intrínsecas. A ideia de um feixe é uma metáfora. Pensem em um feixe de lenha; agora substitua a lenha por propriedades e a amarração do feixe por alguma relação ontológica fundamental, por vezes chamada “copresença”: temos assim um particular concreto. Uma teoria dos feixes pode entender as propriedades que compõem o feixe como universais ou como particulares. Russell, que defendeu de tudo na vida, em algum momento defendeu uma teoria dos feixes que entende as propriedades como universais (1940) – que parece não evitar o problema que os substratos evitam, isto é, o problema sobre a particularidade dos indivíduos. Entender as propriedades como particulares evita

apenas, ou em virtude daquilo que ele é (a massa de um cão, por exemplo), enquanto uma propriedade extrínseca é aquela que um indivíduo tem devido à sua relação com o mundo (o peso do cão).

²⁸Indivíduos que, por terem exatamente os mesmos atributos universais, diríamos que são o mesmo indivíduo. Parece uma consequência dessa teoria admitir essa ideia um tanto quanto estranha.

esse problema e é a estratégia mais comumente utilizada; dois grandes defensores dessa posição são Williams (1953) e Campbell (1990). Ambos defendem que as únicas entidades fundamentais existentes são os tropos – é uma ontologia que postula apenas uma categoria ontológica fundamental.

Como podemos perceber o problema dos universais e o problema dos particulares concretos têm muitas relações entre si. Acredito que esses dois problemas podem ser vistos como os dois principais problemas em ontologia, na medida em que estamos tentando dizer quais são as entidades mais básicas que compõem a realidade. E como podemos perceber, há grande desacordo entre os filósofos na tentativa de oferecer a lista adequada dessas entidades básicas. Muitas vezes esses problemas não são explicitamente distinguidos, e os filósofos os tratam conjuntamente como o problema de determinar as categorias mais gerais e fundamentais da realidade, o que não constitui um erro, mas uma estratégia diferente de abordagem dos problemas.

Seja como for, os problemas dos universais e dos particulares, como qualquer problema filosófico, estão longe de ter uma solução final. Constituem ainda uma estimulante fonte de pesquisa em metafísica. Os filósofos, em geral, depois de determinar quais as entidades mais básicas que admitem, buscam, na sequência, fazer uma análise metafísica das demais noções que “eliminamos”, usando essas entidades básicas. Por exemplo, podemos querer dizer que noções como eventos, mundos possíveis, proposições, leis da natureza, entre muitas outras, são algum tipo entidade composta por aquelas entidades mais básicas. Avaliar o poder das teorias de dar uma solução para aqueles tópicos, nos ajuda a avaliar adequadamente o valor de cada teoria, e o valor das teorias confrontadas entre si.

4. O problema sobre a natureza do tempo

Outro problema muito debatido na metafísica é o problema sobre a natureza do tempo. Desde a antiguidade esse tema tem interessado aos filósofos. Mas a discussão contemporânea sobre o tempo parte de um argumento específico que visa mostrar que o tempo é irreal, i.e., que o tempo não existe. Quem propõe esse argumento é McTaggart (1908), um filósofo muito influenciado por Hegel: o que explica um pouco essa ideia a princípio absurda. Ele propõe uma distinção entre a série A e a série B do tempo. Na primeira parte

do argumento ele procura mostrar que a série A é fundamental para o tempo e, portanto, a série B apenas é insuficiente para a realidade do tempo. Posteriormente, na segunda parte do argumento, procura mostrar que a série A é contraditória – daí que o tempo seja irreal. Muitos filósofos discordam de McTaggart em relação à ideia de que o tempo é irreal, mas concordam com a distinção entre as séries A e B, donde surge a distinção entre teoria A e teoria B. O argumento de McTaggart é relativamente complexo, mas tentarei explicá-lo brevemente aqui focando nos seus pontos principais.

Ele começa o argumento mostrando que há duas maneiras distintas pelas quais podemos considerar as posições no tempo. Temos a série B, na qual cada posição do tempo é *anterior*, *simultânea*, ou *posterior* a alguma das outras posições. E temos a série A, na qual cada posição é *passado*, *presente* e *futuro*. As posições na série B são permanentes, enquanto na série A elas estão em constante mudança. Para vermos como isso se dá, temos apenas que considerar certas atribuições temporais. Por exemplo, o meu nascimento em 1983 é anterior ao nascimento de meu irmão em 1989, e será sempre anterior ao nascimento do meu irmão. Mas o meu nascimento é posterior ao nascimento de minha irmã em 1977, e será sempre anterior ao nascimento dela. Assim, vemos em que sentido as posições na série B são fixas. Entretanto, o meu nascimento foi um dia um futuro muito distante, que foi cada vez mais se aproximando do presente, até que passou a ser presente, e logo começou a se tornar mais e mais passado. Nesse sentido é que as posições na série A não são fixas: os eventos mudam as suas atribuições temporais na série A.

Apesar de haver essas duas séries e elas terem algum tipo de relação, McTaggart argumenta que a série A é mais fundamental que a série B, ou que a série B apenas é insuficiente para constituir o tempo. Isso porque o tempo parece envolver mudança. E a mudança só é possível na série A, de modo que o tempo envolve a série A de uma maneira fundamental.

Essa parte do argumento pode ser colocada do seguinte modo:

1. O tempo envolve mudança.
 2. A mudança só é possível na série A
 3. Se o tempo envolve mudança e a mudança só é possível na série A, então a série A é fundamental para o tempo.
- c. Logo, a série A é fundamental para o tempo.

Alguns filósofos vão contestar (1), que é a ideia de que o tempo envolve mudança (Shoemaker, 1993), outros contestam (2), que é a ideia de que não há mudança na série B (Russell, 1937). Mas precisamos que (1) e (2) sejam conjuntamente verdadeiras para podermos extrair, juntamente com (3), a conclusão de que a série A é fundamental para o tempo. Em geral, os filósofos que contestam a primeira parte do argumento de McTaggart assumem uma teoria B sobre o tempo; teoria essa que toma a série B como suficiente para a existência do tempo. Não entrarei nos detalhes dessa discussão aqui, mas passo a falar sobre a segunda parte do argumento de McTaggart, na qual ele defende que a série A é contraditória.

A segunda parte do argumento é conhecida como o “paradoxo de McTaggart”. Para mostrar que o tempo não existe, McTaggart procura argumentar a favor da ideia de que a série A é contraditória, e que a tentativa de eliminarmos essa contradição leva-nos a uma regressão infinita. Assim, ele afirma que as determinações da série A são incompatíveis entre si. Isso quer dizer que um evento não pode ser passado, presente e futuro. Contudo, McTaggart afirma que os eventos têm todas essas determinações: são passados, presentes e futuros. De modo que chegamos à contradição. Essa segunda parte do argumento pode ser colocada como se segue:

1. Todo evento na série A é passado, presente e futuro.
 2. Nenhum evento pode ser passado, presente e futuro.
 3. Se todo evento na série A é passado presente e futuro e nenhum evento pode ser passado presente e futuro, então a série A é contraditória.
- c. Logo, a série A é contraditória.

Há uma objeção óbvia ao argumento que é antecipada pelo próprio McTaggart. Um evento E não é passado, é presente e é futuro. Suponha que E é um evento presente. Assim, ele *será* passado e *foi* futuro. Se E é um evento futuro, ele *será* presente e passado. Se E é passado, ele *foi* presente e futuro. Dessa forma, não teríamos uma incompatibilidade entre as atribuições temporais, uma vez que os eventos não têm essas atribuições simultaneamente, mas apenas sucessivamente.

McTaggart procura responder a essa objeção do seguinte modo. Para evitarmos a contradição que é colocada inicialmente – E é passado, é presente e é futuro –, invocamos posições de nível secundário da série A – e.g. E é presente, *foi* futuro e *será* passado. Contudo, quando consideramos todas as posições no nível secundário, observamos que há nove posições: é passado, é presente, é futuro, foi passado, foi presente, foi futuro, será passado, será presente, será futuro. McTaggart pensa que todos os eventos ocupam cada uma dessas nove posições da série A. E entre essas posições vai haver aquelas que são incompatíveis, como é presente e é passado. Uma solução para o novo problema é passar para um terceiro nível da série A – e.g. o evento E é presente no presente, é passado no futuro, e é futuro no passado, etc. Esse nível terá vinte e sete posições, ao exaurirmos todas as possibilidades. Mas alguma das combinações possíveis serão novamente incompatíveis, como é presente no passado e é futuro no presente, e todos os eventos ocupam cada uma das posições da série A. Aqui, parece que caímos em uma regressão infinita viciosa, e a única maneira de evitá-la é não dar o primeiro passo. E evitar o primeiro passo é aceitar que a série A é contraditória.

Defensores da teoria A forneceram respostas ao argumento de McTaggart. Um deles, Dummett (1980), em um artigo relativamente complexo, como a própria discussão sobre o tempo, procurou argumentar que McTaggart faz uma suposição duvidosa na segunda parte do argumento contra a irreabilidade do tempo. Segundo Dummett, McTaggart supõe que podemos ter uma descrição completa e consistente da realidade. Mas Dummett argumenta que o que foi mostrado com o argumento da irreabilidade do tempo é que a teoria que toma a série A como fundamental é incompatível com essa suposição. Assim, ao tentarmos especificar todas as verdades da série A, e opondo-as às verdades a partir da perspectiva atual do sujeito, caímos em contradição. Dummett argumenta que, se for o caso da teoria A ser verdadeira, segue-se que é

impossível essa descrição completa da realidade.

Contudo, supondo que essa descrição seja possível, e que dessa forma a série A é contraditória, podemos tentar admitir uma teoria B, que não é contraditória, mas é incompleta, de acordo com o argumento de McTaggart. Uma boa estratégia para um defensor da teoria B é mostrar que de alguma forma é possível a mudança na série B. Nesse sentido, um teórico da série B pode estar de acordo com essa segunda parte do argumento de McTaggart, a favor da ideia de que a série A é contraditória.

Um último detalhe sobre as teorias B e A sobre o tempo é o seguinte. As teorias B são chamadas de “eternistas” ou de “eternalistas”, na medida em que admitem que todos os momentos do tempo são igualmente reais, não havendo uma prioridade ontológica do presente (que é um conceito da teoria A) em relação aos demais momentos. Já na teoria A, teremos algumas variantes, mas a mais comum é o chamado “presentismo”: a ideia de que há uma prioridade ontológica do presente em relação ao passado e ao futuro, ou em outras palavras, apenas o presente é real – o que é bastante intuitivo para nós que vivemos no presente.

5. O problema sobre a persistência dos particulares ao longo do tempo

O problema que veremos agora tem relação com os três primeiros problemas que vimos. Em primeiro lugar, fala sobre particulares concretos, o que envolve os dois primeiros problemas. E também fala sobre o tempo, o que envolve o terceiro problema, aquele sobre a natureza do tempo. Como veremos, dependendo da concepção que temos sobre o tempo, isso vai influenciar de uma maneira crucial na nossa concepção sobre o que é para um particular persistir ao longo do tempo. Veremos que essas duas concepções implicam concepções de persistência ao longo do tempo completamente diferentes.

Para começar, os particulares concretos são entidades ditas “espaçiotemporais”. Isso quer dizer que estão no espaço e no tempo. Estar no espaço é algo relativamente fácil de entender e não precisamos entrar em muitos detalhes: os particulares concretos são entidades feitas de partes e suas partes estão dispostas nas três dimensões do espaço. Mas o que é estar no tempo? Sabemos que os particulares concretos, pelo menos os que estamos mais familiarizados, surgem em um momento do tempo, desenvolvem-se para, por fim,

perecerem em algum momento posterior. Pensemos em Sócrates. Havia um tempo em que ele não existia, até que foi concebido por seus progenitores. Viveu por muitos anos e veio a falecer por volta de seus 70 anos. Quando dizemos que um particular como Sócrates persiste no tempo, queremos dizer que há algum tipo de identidade entre o bebê Sócrates e o Sócrates idoso, eles são as mesmas pessoas. E o problema sobre a persistência do tempo é justamente esse de explicar como pode haver essa identidade entre coisas que parecem tão diversas.

Temos dois tipos de teorias que buscam explicar o fenômeno da persistência ao longo do tempo, são elas o endurantismo e o perdurantismo²⁹. Essas são duas formas muito diversas de se considerar a persistência. A diferença fundamental entre esses dois tipos de teorias é que, para o endurantismo, os particulares concretos persistem no tempo na medida em que existe total e completamente em momentos distintos do tempo; enquanto para o perdurantismo, os particulares concretos são agregados de suas partes temporais, e cada uma dessas partes existe em um momento do tempo apenas. Isso quer dizer que, contrário ao endurantismo, para o perdurantismo um mesmo particular concreto não existe total e completamente em momentos distintos do tempo; a cada momento do tempo temos uma parte temporal do que chamamos “o particular concreto”.

Retomando o exemplo de Sócrates, na perspectiva endurantista, o Sócrates bebê, o Sócrates adulto e o Sócrates idoso são uma e a mesma pessoa que ocupa momentos diferentes do tempo. Já na perspectiva perdurantista, para cada um desses estágios diferentes da vida de Sócrates temos uma parte temporal diferente em diferentes momentos. Estritamente falando, para o perdurantismo, o Sócrates bebê e o Sócrates idoso não são a mesma entidade, não há uma identidade numérica³⁰ entre eles. Aliás, há várias partes temporais do Sócrates bebê, e uma discussão importante seria para

²⁹Essa terminologia se deve a Lewis (1986). Essas palavras veem do inglês “endurantism” e “perdurantism”, respectivamente.

³⁰Uma entidade é numericamente idêntica a outra, se elas são exatamente a mesma entidade. Exemplo: Pelé é idêntico a Edson Arantes. Há também a noção de identidade qualitativa: é o sentido em que gêmeos univitelinos são idênticos – não são a mesma entidade, mas entidades qualitativamente muito semelhantes.

identificar qual é a duração de uma, e apenas uma, parte temporal.

Um detalhe importante na diferença entre o endurantismo e o perdurantismo é o fato de que o endurantismo³¹, em geral, está associado a uma teoria A do tempo, em especial ao presentismo – a ideia de que apenas o presente é real. Essa posição é também conhecida como tridimensionalismo, dado que supõem que a realidade compreende apenas as três dimensões espaciais. Por outro lado, o perdurantismo³² está associado frequentemente a uma teoria B sobre o tempo, ao chamado eternalismo – a ideia de que todos os momentos do tempo são igualmente reais, e constitui apenas mais uma dimensão, juntamente com as dimensões espaciais. Essa posição é conhecida por quadrimensionalismo, isso porque adiciona às dimensões espaciais uma dimensão temporal.

Apesar de ser um caminho natural para um teórico da teoria B sobre o tempo ser também um perdurantista e, por outro lado, um teórico da teoria A ser também endurantista, essas posições podem não estar essencialmente ligadas, admitindo teorias cruzadas. Há filósofos, por exemplo Mellor (1981), que pensam que uma teoria B é compatível com o endurantismo. Esses aceitam uma teoria B sobre o tempo, mas rejeitam o perdurantismo e sua noção central de partes temporais.

A compatibilidade de um certo tipo de teoria A com o perdurantismo é mais óbvia. Uma teoria A que não seja presentista parece facilmente adaptável ao perdurantismo. Temos tipos de teorias A que admitem, não apenas o presente como existente, mas também o passado e o futuro. Uma versão dessa teoria encontramos em Broad (1923). A diferença para uma teoria B nesse caso é que nessa teoria A os particulares estão constantemente mudando em relação às atribuições temporais, enquanto na teoria B os particulares têm uma posição fixa na linha do tempo. Todavia a mais popular das teorias A, o presentismo, é incompatível com o perdurantismo.

Uma vantagem do endurantismo associado ao presentismo é o seu

³¹Entre os filósofos que defendem o endurantismo encontramos Geach (1966; 1967), Chisholm (1976), Lowe (1983; 1998), Simons (1987), van Inwagen (1990a; 1990b), Fine (2008/2006), e Shoemaker (2015).

³²Entre os filósofos que defendem o perdurantismo, temos Quine (1950; 1960), Taylor (1955), Lewis (1971; 1976; 1986), e Armstrong (1980).

apelo intuitivo inicial. Parece uma concepção do senso comum a ideia de que o presente é o único momento realmente existente; talvez por estarmos nele, em contato relativamente direto. O perdurantismo, por outro lado, tem o problema de ser contraintuitivo, pelo menos desse ponto de vista do senso comum. Nesse sentido, pelo menos a princípio, o ônus da prova é do perdurantismo, e os filósofos defensores dessa posição buscam oferecer argumentos a favor dela. Duas linhas de argumentação são frequentes. Na primeira delas, os perdurantistas buscam argumentar a favor da ideia de que, apenas nessa perspectiva, podemos dar uma caracterização adequada da mudança de propriedades nos particulares. Na segunda, buscam argumentar a favor da ideia de que, apenas nessa perspectiva, podemos compreender adequadamente a mudança nas partes dos particulares. No debate, os endurantistas buscam contestar de alguma forma esses argumentos. Considerarei aqui de maneira breve apenas a primeira linha de argumentação.

Antes de entrar no argumento sobre a mudança nas propriedades, consideremos muito brevemente um motivo para preferirmos o perdurantismo. Argumenta-se que o perdurantismo está mais de acordo com a nossa visão científica do mundo. Supostamente, a teoria da relatividade se adequa bem ao perdurantismo, na medida em que toma o tempo como mais uma dimensão da realidade. Entretanto há endurantistas que parecem não estar de acordo com isso. Pensam que podemos dar uma descrição teórica adequada da realidade a partir do endurantismo.

Entrando no argumento sobre a mudança nas propriedades dos particulares concretos, um dado básico sobre as nossas crenças sobre o mundo é a ideia de que os particulares mudam ao longo do tempo, e que eles persistem mesmo com essa mudança. Os perdurantistas buscam argumentar que o endurantismo não tem como lidar com esse fato básico sobre as nossas crenças sobre o mundo. O ponto é o seguinte: a persistência parece envolver a identidade numérica de uma coisa existindo em dois momentos distintos. Mas os perdurantistas argumentam que há um problema lógico com essa suposição. Ela contraria um princípio amplamente aceito: a indiscernibilidade dos idênticos. Esse princípio diz que se dois objetos são idênticos, então eles têm exatamente as mesmas propriedades. Por exemplo, se digo que Pelé é jogador de futebol, e eu sei que Pelé e Edson Arantes são uma só pessoa, posso concluir que Edson Arantes é jogador de futebol.

Mas retomemos nosso exemplo predileto: Sócrates. Imaginemos que Sócrates tenha passado uma temporada em algum lugar no norte da Europa, que sabemos ser bem frio. Imaginemos também, que depois dessa temporada ele retorne a Atenas bastante pálido pela falta de sol, mas logo ganha novamente uma cor morena. Diríamos, dada a verdade do princípio da indiscernibilidade dos idênticos, que o Sócrates pálido da temporada em questão e o Sócrates moreno da temporada seguinte não são a mesma pessoa dado que, se fossem, teriam que ter exatamente as mesmas propriedades. Mas eles diferem, pelo menos, no que diz respeito às propriedades de ser pálido e de ser moreno. Disso, o perdurantista conclui que a identidade que há entre o Sócrates de uma temporada e o Sócrates da temporada seguinte não pode ser uma identidade numérica. E isso seria um motivo para admitirmos a existência de partes temporais, ao invés da teoria endurantista sobre a persistência dos particulares concretos. A ideia é então que temos partes temporais de Sócrates diferentes sendo consideradas, mas o Sócrates é um agregado de todas essas partes.

Uma forma de responder a isso seria dizer que a lógica que estamos considerando é não temporal. Nesse sentido, não podemos querer que o princípio em causa seja uma verdade, se incluirmos aqui índices temporais – o princípio já não seria verdadeiro. Isso porque parece um dado básico sobre a realidade que os particulares mudam pelo menos algumas de suas propriedades ao longo do tempo, sendo, ainda assim, o mesmo particular. Seria um pouco absurdo eu ter um carro branco em um momento t_1 e em um momento t_2 ele ter sido pintado de vermelho, e a partir disso eu concluir que são carros diferentes que tenho nos momentos t_1 e t_2 . Parece óbvio que é o mesmo carro com cores diferentes. O endurantista talvez tenha que recorrer a uma distinção entre os atributos que compõem o particular concreto, mostrando que certos atributos são essenciais ao particular, enquanto outros não são essenciais e o particular pode os perder à medida em que ganha outros. Assim, poderiam dizer que Sócrates não pode deixar de ser humano (pertencer à espécie) sem deixar de existir como Sócrates, mas a sua cor pálida, ao voltar do norte, pode ser perdida sem prejuízos ao particular.

Conclusão

Chegamos ao fim, sem uma definição explícita e precisa sobre o que é a metafísica. Defini-la adequadamente é algo difícil, como é difícil definir outras disciplinas, como a matemática e a física. Mas espero ter conseguido indicar minimamente o que os metafísicos fizeram e ainda fazem através da longa história dessa disciplina. Na sequência indico alguns livros que podem ser interessantes para aqueles que gostariam de adentrar mais nesse terreno fértil que é a metafísica.

Questões para revisão

1. De que maneira Aristóteles caracteriza a metafísica?
2. De que maneira os racionalistas modernos compreendiam a metafísica?
3. Quais são as principais posições sobre o problema dos universais?
4. Quais são as principais posições sobre o problema dos particulares concretos?
5. Explique a distinção entre teoria A e teoria B sobre o tempo.
6. Explique a distinção entre endurantismo e perdurantismo e o modo como compreendem a persistência dos particulares concretos ao longo do tempo.
7. Explique o argumento da mudança nas propriedades a favor do perdurantismo.

Questões para discussão

1. A posição de Kant sobre a possibilidade da metafísica é coerente?
2. A tese relativista pode ser racionalmente defendida?
3. E sobre o cientificismo, há alguma forma de defendê-lo racionalmente?
4. Sobre o problema dos universais, qual vertente lhe parece mais intuitiva, realismo ou nominalismo? E por quê?
5. Dentre as concepções de particulares concretos, qual lhe parece mais defensável?
6. É possível haver mudança na série B?
7. A série A é realmente contraditória?

Leituras recomendadas

- CID, Rodrigo (2019). *Leis da Natureza: Uma Abordagem Filosófica*. Macapá: Editora da UNIFAP. É um bom livro para aqueles que se interessam pelo tema da causalidade, leis da natureza e a relação desses tópicos com o problema dos universais.
- GARRETT, B., (2008). *Metafísica: conceitos básicos em filosofia*. São Paulo: Artmed. É um bom livro introdutório sobre o problema dos universais, o problema sobre os particulares e o problema sobre o tempo, entre outros tópicos em metafísica.
- LOUX, M., (2006). *Metaphysics: a contemporary introduction*, 3ª Ed. New York: Routledge. Apesar de estar em inglês, o livro é uma das melhores introduções à metafísica. Trata sobre todos esses tópicos em metafísica mencionados, entre outros.
- LOWE, E.J., (2008). “A Possibilidade da Metafísica”. *Crítica na rede*. <https://criticanarede.com/posmetafisica.html>. Esse é um bom texto introdutório sobre o que é a metafísica e a possibilidade da metafísica.
- SCHWARTZ, S. P., (2017). *Uma breve História da Filosofia Analítica: de Russell a Rawls*. São Paulo: Edições Loyola. O livro introduz trata de alguns tópicos importantes sobre a metafísica na filosofia analítica; em especial há um capítulo sobre o renascimento da metafísica e um capítulo sobre as modalidades metafísica (necessidade, possibilidade e contingência).

Referências

- ARISTÓTELES. *Metafísica*. Tradução Marcelo Perine. São Paulo: Edições Loyola, 2002.
- _____. *Categorias*. Trad., introd. e notas de José Veríssimo Teixeira da Mata. Goiânia: Ed. UFG; Alternativa, 2005.
- ARMSTRONG, D. M. (1978a). *Nominalism and Realism, vol. 1 of Universals and Scientific Realism*. Cambridge: Cambridge University Press.
- _____. (1978b). *A Theory of Universals, vol. 2 of Universals and Scientific Realism*. Cambridge: Cambridge University Press.
- _____. (1980). “Identity Through Time”, em Peter van Inwagen (ed.) *Time and Cause*, Dordrecht: D. Reidel: 67–78.

- _____. (1989). *Universals: an opinionated introduction*. Boulder: Westview Press.
- BROAD, C.D. (1923). *Scientific Thought*, London: Kegan Paul.
- CAMPBELL, K. (1990). *Abstract Particulars*. Oxford: Basil Blackwell.
- CHISHOLM, R. M. (1976). *Person and Object*. London: Allen and Unwin.
- DUMMETT, M. (1980). "A defense of McTaggart's proof of the unreality of time". *Truth and other Enigmas*. Cambridge, Mass.: Harvard University Press, 351-357.
- FINE, K. (2008/2006). "In Defence of Three-Dimensionalism", *Philosophy*, 83, supplement 62 (2008): 1–16; publicado pela primeira vez como 'Defense' no *Journal of Philosophy* CIII.12 (2006): 699–714.
- GEACH, P.T. (1966). "Some Problems about Time". *Proceedings of the British Academy*, 60.
- _____. (1967). "Identity". *Review of Metaphysics*, 21: 3–12.
- HEIL, J. (2003). *From an Ontological Point of View*. Oxford: Oxford University Press.
- HUME, D. (1748). *Investigações sobre o Entendimento Humano e sobre os princípios da moral*. Tradução José Oscar de Almeida Marques. São Paulo: Editora Unesp, 2004.
- KANT, I. (1787). *Crítica da Razão Pura*. Tradução: Manuela Pinto dos Santos e Alexandre Fredrique Morujão. Lisboa: Fundação Calouste Gulbenkian, 2001.
- KRIPKE, S. (1980). *O Nomear e a Necessidade*. Tradução Ricardo Santos e Teresa Filipe. Lisboa: Gradiva, 2012.
- LEWIS, D. (1971). "Counterparts of Persons and their Bodies", *Journal of Philosophy*, 68: 203–211.
- _____. (1983). "New Work for a Theory of Universals". *Australasian Journal of Philosophy*, 61, 4, pp. 343-77.
- _____. (1986). *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- LOUX, Michael. (1978). *Substance and Attribute*. Dordrecht: D. Reidel Publishing Company.
- LOWE, E. J. (1983). "Instantiation, Identity and Constitution". *Philosophical Studies*, 44: 45–59.
- _____. (1998). *The Possibility of Metaphysics*. Oxford: Oxford University Press.
- _____. (2006). *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Oxford University Press.
- MARTIN, C.B., (1980). "Substance Substantiated". *Australasian Journal of*

- Philosophy* 58 (1):3 – 10.
- MCTAGGART, J. M. E. (1908). “A Irrealidade do Tempo”. *Kriterion*. Trad. César Schirmer dos Santos, nº 130, 2014, p. 747-764.
- MELLOR, D.H. (1981). *Real Time*. Cambridge: Cambridge University Press.
- PLATÃO. *Parmênides*. Tradução Maura Iglesias e Fernando Rodrigues. Rio de Janeiro: Ed. PUC-Rio, São Paulo Edições Loyola, 2003.
- QUINE, W.V.O. (1948). “Sobre o Que Há”. Em: W. V. O. Quine. *De um Ponto de Vista Lógico: nove ensaios lógico-filosóficos*. Tradução A. I. SEGATTO. São Paulo: UNESP, 2011.
- _____. (1950). “Identidade, Ostensão e Hipóstase”. Em: W. V. O. Quine. *De um Ponto de Vista Lógico: nove ensaios lógico-filosóficos*. Tradução A. I. SEGATTO. São Paulo: UNESP, 2011.
- _____. (1960). *Palavra e Objeto*. São Paulo: Editora Vozes, 2010.
- RODRIGUEZ-PEREYRA, G. (2000). “What is the Problem of Universals?” *Mind*, 109, 434, pp. 255-73.
- _____. (2002) *Resemblance Nominalism*. Oxford: Oxford University Press.
- RUSSELL, B. (1912). *Os Problemas da Filosofia*. Tradução, introdução e notas de Desidério Murcho, Lisboa: Edições 70, 2008.
- _____. (1927). *An Outline of Philosophy*. London: George Allen and Unwin.
- _____. (1940). *An Inquiry into Meaning and Truth*. New York: W. W. Norton.
- SHOEMAKER, S. (1993). “Time without Change”. *The Philosophy of time*. Ed. LE POIDEVIN, R., MACBEATH, M. Oxford: Oxford University Press.
- _____. (2015). “Persistence and Properties”. *Journal of the American Philosophical Association*, 1.3: 433–48.
- SIMONS, P. (1987). *Parts: A Study in Ontology*. Oxford: Clarendon Press.
- SPADE, P. V. (1994). *Five Texts on the Mediaeval Problem of Universals*. Cambridge: Hackett Publishing Company.
- STOUT, G.F. (1923). “Are the Characteristics of Particular Things Universals or Particular?” *The Problems of Universals*. Ed. Charles Landesman, New York: Basic Books, 1971.
- STRAWSON. P.F. (1959). *Individuals*. London: Methuen.
- TAYLOR, R. (1955). “Spatial and Temporal Analogues and the Concept of Identity”. *Journal of Philosophy*, 52, 599–612.
- TOOLEY, Michael. (1977) The Nature of Laws. *Canadian Journal of Philosophy*, Vol. VII, nº 4.

_____. (1987) *Causation*. Oxford: Clarendon Press.

VAN INWAGEN, P. (1990). "Four-Dimensional Objects". *Noûs*, 24: 245–55.

WILLIAMS, D. C. (1953) The Elements of Being. *Rev. Metaphysics*. 7. 3-18, 171-192.

3

Epistemologia

Delvair Custódio Moreira

Neste trabalho eu pretendo apresentar algumas questões importantes da epistemologia tradicional e da relativamente recente epistemologia social. Por “epistemologia tradicional” eu quero me referir a área da filosofia que se ocupa com questões relativas às condições do conhecimento e as normas que regem a racionalidade ou justificação de crenças. A epistemologia social – uma ramificação da epistemologia tradicional – ocupa-se com questões semelhantes, porém, levando em conta as diversas interações sociais nas quais os indivíduos estão inseridos. Tipicamente, a epistemologia tradicional ignorou os elementos sociais envolvidos nos processos de formação de crenças e obtenção de conhecimento por pressupor que o conhecimento ou os padrões de racionalidade e justificação de crenças são alcançados por sujeitos individuais empregando suas próprias capacidades cognitivas. O surgimento da epistemologia social deve-se em muito a críticas contra a esse pressuposto individualismo epistêmico da epistemologia tradicional. Por essa razão, escolhi apresentar alguns dos problemas trabalhados em epistemologia tradicional que evidenciam individualismo epistêmico, a fim de tornar claras as objeções ao individualismo e os problemas investigados em epistemologia social. A abordagem de epistemologia social aqui apresentada, no entanto, não deve ser entendida como uma tentativa de reformulação da epistemologia tradicional, mas como uma proposta complementar a essa.

1. Análise do conhecimento

Uma importante questão em epistemologia diz respeito à análise do conhecimento. O tipo de conhecimento que epistemólogos tipicamente estão interessados é o conhecimento sobre fatos, ou, como é tecnicamente chamado,

“conhecimento proposicional”. O conhecimento proposicional é aquele ao qual nos referimos em expressões do tipo “S sabe que p ” (onde “S” é um sujeito qualquer e “ p ” é uma proposição qualquer), como em “João sabe que a água é H_2O ” – que se distingue do conhecimento por habilidade, que nos referimos em expressões do tipo “S sabe como”, como em “João sabe como tocar violão”; e do conhecimento por familiaridade, que nos referimos em expressões como “S conhece x ”, como em “João conhece o litoral Norte”. A análise do conhecimento proposicional visa responder a seguinte pergunta: quais são as condições necessárias e suficientes para ser o caso que S saiba que p ? Uma influente resposta a essa questão, que remonta à Platão, nos diz que conhecimento é crença verdadeira justificada. Para essa “análise clássica” (como podemos chamar) de conhecimento S sabe que p quando as seguintes condições são satisfeitas: (i) S acredita que p ; (ii) é verdade que p (iii) S está justificado em acreditar que p . Digamos que João saiba que há dez latas de cerveja no congelador. Segundo a análise clássica, tripartite, de conhecimento isto significa que João *acredita* que há dez latas de cerveja no congelador, e está *justificado* em acreditar nisso (talvez porque se lembre que comprou dez latas de cerveja e as colocou há pouco no congelador), e é *verdade* que as dez latas estão no congelador. Se alguma dessas três condições não forem satisfeitas, então não é o caso que João saiba que há dez latas de cerveja no congelador. Imagine que João *não acredite* que há dez latas de cerveja no congelador, porque pensa erroneamente que seu irmão bebeu duas latas. Não parece correto, agora, dizer que João sabe que há dez latas de cerveja no congelador, mesmo que isso continue sendo verdade. Suponha agora que João acredite que há dez latas de cerveja no congelador, porque se lembra de ter comprado as dez latas e as colocado no congelador. Mas, sem que João percebesse, seu irmão bebeu duas latas. Novamente, não se pode mais dizer que João sabe que há dez latas de cerveja no congelador, uma vez que agora há oito latas (ou seja, agora é falso que há dez latas de cerveja no congelador). Por fim, considere outra variação no exemplo: quem comprou as dez latas de cerveja foi o irmão de João e as guardou no congelador sem João ver. O irmão de João faz uma brincadeira e lhe diz “se você adivinhar quantas latas de cerveja tem no congelador, eu lhe dou metade”. Sem pestanejar, João responde “tem dez latas de cerveja no congelador!”. Mais uma vez, não parece ser o caso que João saiba quantas latas de cerveja há no congelador, pois, uma vez que foi um palpite de sorte, João *não*

está justificado em acreditar nisso.

Quadro 1: Condições necessárias e suficientes para o conhecimento e o Problema de Gettier

A análise clássica de conhecimento é uma tentativa de estabelecer as condições necessárias e suficientes para o conhecimento, mas o que isso significa? Diz-se que uma condição *c* é *necessária* para *x* ser *F* quando tudo que é *F* satisfaz *c*, mas nem tudo que satisfaz *c* é *F*. Por exemplo, uma condição necessária para alguém ser mineiro é ser brasileiro, mas ser brasileiro não faz de alguém um mineiro. Já uma condição *c* é *suficiente* para *x* ser *F* quando tudo que satisfaz *c* é *F*, mas nem tudo que é *F* satisfaz *c*. Por exemplo, ser mineiro é uma condição suficiente para ser brasileiro, mas alguém pode ser brasileiro sem ser mineiro. Quando uma condição *c* é necessária e suficiente para *x* ser *F* isso significa tudo que satisfaz *c* é *F* e todo *F* satisfaz *c*. Uma condição necessária e suficiente para alguém ser mineiro é ter nascido no estado de Minas Gerais. Do mesmo modo, a análise clássica de conhecimento nos diz que as condições necessárias e suficientes para alguém ter conhecimento que *p* é ter uma crença verdadeira justificada que *p*, o que significa que toda vez que *S* tem uma crença verdadeira justificada que *p*, então *S* sabe que *p* e toda vez que *S* sabe *p*, então *S* tem uma crença verdadeira justificada que *p*.

Durante muito tempo a análise clássica do conhecimento foi considerada bem sucedida. No entanto, em 1963, o epistemólogo Edmund Gettier em um artigo intitulado “É crença verdadeira justificada conhecimento?” ofereceu uma série de contraexemplos à análise clássica que mostravam que as três condições não eram suficientes para o conhecimento. Ou seja, o “Problema de Gettier”, como ficou conhecido, mostra que há casos onde *S* tem uma crença verdadeira justificada que *p* e, apesar disso, *S* não sabe que *p*. Um exemplo estilo Gettier seria o seguinte: João tem ingressos para a sessão das cinco no cinema local e decide sair de casa as quatro horas para não chegar atrasado. As três horas, João olha confere em seu relógio de pulso e se certifica que ainda são três horas. O que João não sabe é que seu relógio parou exatamente as três horas no dia anterior e que por um acaso ele conferiu as horas exatamente quando eram três horas hoje. Neste exemplo, João tem uma crença verdadeira justificada que são três horas, mas isso não é conhecimento: foi por um acaso que ele olhou justamente na hora que o relógio parado estaria certo naquele dia.

Desde a publicação do artigo de Gettier é praticamente consensual que crença, verdade e justificação não são condições *suficientes* para o conhecimento. Porém,

para muitos epistemólogos essas condições ainda são *necessárias*. Por isso, em muitas discussões epistemológicas é pressuposto que conhecimento é *peelo menos* crença, verdadeira justificada. Ainda que a questão sobre quais são as condições necessárias e suficientes para o conhecimento ainda seja uma questão em aberto para epistemólogos.

2. Crenças e Justificação Epistêmica

Central na análise clássica de conhecimento é a noção de crença justificada. Uma crença é uma atitude doxástica, isto é, uma atitude que temos com relação ao valor de verdade de uma proposição. Por exemplo, considere a seguinte proposição: “há vida inteligente fora da Terra”. O leitor pode *acreditar* na verdade desta proposição, pode *desacreditar* (o que equivale a acreditar que a proposição é falsa), ou ainda pode julgar que não tem informação o suficiente nem para acreditar que a proposição é verdadeira, nem que a proposição é falsa, o que consiste em *suspender o juízo* com relação à proposição. Atitudes como crença e descrença podem variar em grau. Talvez o leitor acredite com maior grau que há vida fora da Terra, mas tenha menor convicção de que há vida *inteligente* fora da Terra. Porém, considerações de grau ou força da crença geralmente, por razões didáticas, não são levadas em conta na análise clássica de conhecimento. Em alguns casos consideramos correto ou apropriado ter determinada atitude doxástica em vez de outra. Por exemplo, suponha que alguém, após uma breve busca, não encontre o controle de sua TV e então passe a acreditar que o controle foi roubado. Seria essa uma crença adequada? Certamente o leitor diria que não. Costumamos usar termos avaliativos para dizer se uma atitude doxástica é ou não adequada em uma determinada situação. No exemplo anterior poderíamos dizer que é irracional ou infundada a crença de que o controle foi roubado. O termo técnico tipicamente utilizado por epistemólogos para avaliar uma crença é “justificação”: diz-se que uma crença é justificada para um sujeito quando é apropriado tomar essa atitude doxástica e injustificada quando não é. Um ponto importante a se considerar é que quando falamos em justificação, em epistemologia, estamos falando em justificação *epistêmica*. Isso porque pode ser adequado, em algumas situações, manter uma crença considerando algum benefício prático, mas não necessariamente epistêmico. Por exemplo, suponha um jogador de futebol errou a maioria dos

pênaltis que cobrou no campeonato. Estatisticamente parece razoável que ele acredite que irá errar o próximo pênalti que for cobrar. Mas jogadores podem ter melhor desempenho se forem mais confiantes, então pode ser adequado do ponto de vista prático para este jogador acreditar que vai marcar um gol no próximo pênalti que cobrar. A justificação epistêmica, por outro lado, é considerada quando temos como meta acreditar em verdades e evitar acreditar em falsidades. Voltemos ao exemplo do controle remoto: o que há de errado com a crença do sujeito que acredita que o controle remoto foi roubado, após não encontrá-lo, é que o mero desaparecimento do controle não parece indicar que seja verdade que o controle foi roubado – é muito mais razoável pensar que ele está perdido em algum lugar, dado que é comum perder controles de TV.

2.1. Evidencialismo

A questão sobre o que torna uma crença justificada é tema de grande controvérsia entre os epistemólogos. Segundo uma linha argumentativa uma crença está justificada para um sujeito quanto este tem evidência de que a crença é verdadeira. O termo “evidência” aqui não deve ser confundido com um sentido material de evidência. Por exemplo, o DNA de alguém encontrado numa cena de crime pode ser evidência de que este é o culpado. Mas não é esse o sentido material de “evidência” que epistemólogos têm em mente quando dizem que estar justificado em acreditar que p é estar em posse de evidência de que p . Evidência no sentido epistêmico é um certo estado mental no qual o indivíduo se encontra ou tem acesso (no sentido de estar ciente) que dê suporte à verdade da crença para o sujeito. Geralmente, a noção de evidência é entendida também como uma crença. Por exemplo, a minha crença de que as chaves do portão estão guardadas na gaveta do armário é justificada porque eu acredito tê-las guardado no armário, quando cheguei em casa.

O leitor pode já ter notado que temos um potencial problema aqui para o evidencialista: se o que torna uma crença C justificada é uma crença D , então parece que D também tem de estar justificada, a fim de que C esteja justificada. A justificação da crença D pode vir de uma crença E , mas E também precisaria estar justificada e assim por diante. Isso nos conduziria a um regresso infinito na cadeia de justificação o que constitui em um sério problema para a teoria evidencialista da justificação epistêmica. A razão disso é que, alegam alguns

epistemólogos, ninguém pode conter um número infinito de crenças na mente e, se assim for, ninguém estaria justificado em acreditar em coisa alguma. Talvez a cadeia de justificação retorne ao ponto inicial, formando um círculo: C é justificado por D, D é justificado por E e E é justificado por C. Isto também seria um problema para o evidencialista: imagine João acredita que o Brasil tem a melhor seleção de futebol do mundo. João acredita o Brasil tem a melhor seleção de futebol do mundo, porque acredita que a seleção brasileira de futebol irá ganhar a próxima copa do mundo e João acredita que a seleção brasileira irá ganhar a próxima copa do mundo, porque o Brasil tem a melhor seleção de futebol do mundo. Como fica claro, João não nos deu razão alguma que justifique sua crença inicial. Se a cadeia de justificação for circular, portanto, isso implica que não há justificação alguma.

Por fim, podemos considerar que a cadeia de justificação termine arbitrariamente em uma crença qualquer. Isso também não seria satisfatório, pois, como ocorre nas duas primeiras alternativas, não parece que alguém estaria justificado em acreditar em C, se decidir parar *arbitrariamente* de oferecer razões que sustentem C: seria como se respondêssemos “porque sim!” quando fossemos questionados sobre o porquê de acreditarmos em algo. Desta forma, se a justificação funciona como o evidencialismo sugere, parece impossível que alguém esteja justificado em acreditar em qualquer coisa que seja, pois a cadeia de justificação para uma crença ou seria infinita, ou seria circular, ou seria interrompida arbitrariamente. Esse problema é conhecido na literatura como o “Trilema de Agripa” e constitui-se em um sério desafio para a teoria do conhecimento, uma vez que a consequência do Trilema é o ceticismo. Uma vez que consideramos que conhecimento é, pelo menos, crença verdadeira justificada, se não é possível para alguém estar justificado em acreditar no que quer que seja, então não é possível conhecer o que quer que seja. No entanto o evidencialista tem duas estratégias para solucionar esse problema imposto pelo Trilema de Agripa. Essas estratégias dizem respeito à *estrutura da justificação* e é o cerne do debate entre coerentistas e fundacionistas sobre a justificação epistêmica.

2.2. A Estrutura da Justificação: Fundacionismo e Coerentismo

O fundacionismo defende que há crenças especiais na cadeia de justificação que são chamadas de “crenças básicas”. Tais crenças fornecem justificação para as outras crenças não-básicas, sem que elas mesmas precisem ser justificadas *por outras crenças*. Se isto estiver correto há uma alternativa não-cética ao Trilema de Agripa: o regresso na cadeia de justificação termina em uma crença básica que é justificada sem recurso a outras crenças. No entanto, para que essa resposta funcione é preciso explicar o que torna uma crença básica justificada. Há dois tipos de fundacionismos, que fornecem explicações diferentes para a questão anterior. O primeiro, chamado “fundacionismo cartesiano”, afirma que as crenças básicas têm uma propriedade especial que as autojustifica e o segundo, chamado “fundacionismo empírico”, afirma que as crenças básicas derivam sua justificação da experiência sensorial. Começemos com o fundacionismo cartesiano. Segundo o fundacionista cartesiano, crenças básicas são crenças sobre nossos estados mentais. Por exemplo: estar ansioso, estar com sede, estar com dores, são exemplos de estados mentais. Um sujeito pode formar, por introspecção, crenças sobre os estados mentais no qual se encontra: S acredita que está ansioso, S acredita que está com sede, S acredita que está com dores, e assim por diante. Mas o que há de especial com crenças desse tipo? Tipicamente, fundacionistas cartesianos argumentam que crenças sobre estados mentais (ou crenças introspectivas) são *indubitáveis* ou *infalíveis*. Por “indubitável” entenda-se que não podemos duvidar do conteúdo dessas crenças. Por exemplo, alguém que está com dores não pode duvidar de que está sentido dores. Mas seria razoável pensar que do fato de que não podemos duvidar do conteúdo de uma crença nós estamos justificados em acreditar nela sem outras evidências? Alguns epistemólogos argumentam que a indubitabilidade é uma limitação psicológica que não deveria ter qualquer valor epistêmico. Por exemplo, Goldman (1979) argumenta que “um fanático religioso pode ser psicologicamente incapaz de duvidar dos princípios da sua fé, mas isso não justifica a sua crença neles”. Portanto, a indubitabilidade não parece ser uma boa candidata à propriedade autojustificante de crenças básicas. A infalibilidade de crenças sobre estados mentais, no entanto, parece ser uma propriedade capaz de autojustificar as crenças deste tipo. Dizer que uma crença sobre um estado mental é infalível significa que não podemos estar errados sobre elas. Por

exemplo, se estou sentido dores de cabeça, eu não posso estar errado sobre isso; se sinto sede, não tem como eu estar enganado sobre isso; e assim por diante. Ora, se eu não posso estar errado sobre o conteúdo de uma crença, parece razoável aceitar que eu não preciso de qualquer outra evidência ou crença para estar justificado em ter essa crença. Mas crenças introspectivas parecem dizer respeito apenas sobre o mundo interior de nossas mentes e isso parece muito pouco. Crenças básicas deveriam também fornecer justificção para outras crenças, como, por exemplo, para as crenças que temos sobre o mundo externo a nós. Há, no entanto, uma classe de crenças sobre estados mentais que ainda não investigamos e que contariam como crenças básicas para o fundacionista cartesiano: as chamadas “crenças de aparência”, ou seja, crenças sobre como as coisas aparecem para nós. Quando estamos vendo um objeto temos uma experiência sensorial (ou perceptiva) do objeto em nossa mente. Essa experiência é um estado mental sobre o qual nós podemos formar crenças. E assim como não podemos estar enganados sobre estar com sede ou com dores, também não nos enganamos sobre ter algum tipo de experiência perceptiva: tenho uma experiência perceptiva de que a caneca sobre a mesa é vermelha e não posso estar enganado acerca dessa experiência. Posso, evidentemente, estar errado sobre haver uma caneca vermelha sobre a mesa, digamos, se eu estiver alucinando. Mas não posso estar errado sobre o fato de que *me parece* haver uma caneca vermelha ao lado de meu notebook neste momento. Sobre a infalibilidade de crenças de aparência, Santo Agostinho, em *Contra os Acadêmicos*, escreve:

Pois não vejo como o Acadêmico possa refutar alguém que diz: sei que isso me parece branco, sei que isso deleita meus ouvidos, sei que este odor me agrada, sei que aquilo tem gosto doce, sei que aquilo é frio para mim.

– Mas, dize-me antes se são amargas em si mesmas as folhas do oleastro que tanto apetezem à cabra!

– Ó homem petulante! Não é mais modesta a cabra? Não sei o que as folhas são para o animal. Para mim são amargas. O que mais queres?

– Mas talvez também há algum homem para quem não são amargas.

– Queres cansar-me? Por acaso eu disse que são amargas para todos? Disse que são amargas para mim, e não afirmo que isso é sempre assim. Não acontece que por uma causa ou outra, a mesma coisa uma vez tem gosto doce, outra vez amargo? Afirmo o seguinte: quando um homem saboreia alguma coisa, pode jurar de boa-fé que sabe que tal coisa é suave ou não ao seu paladar e não há sofisma grego que possa retirar-lhe esse conhecimento (p. 78).

Muito bem. Suponha que o fundacionista cartesiano esteja certo sobre a infalibilidade das crenças acerca de estados mentais. A questão que se coloca agora é como, a partir dessas crenças, nós podemos ganhar justificação para as crenças não-básicas. Em particular, para crenças sobre objetos no mundo externo. Como a crença de que me parece haver uma caneca vermelha sobre a mesa justifica a crença de que há uma caneca vermelha sobre a mesa? Afinal, de tudo que foi dito, podemos assumir apenas que não estou errado quanto a como as coisas me parecem, mas ainda posso estar enganado sobre como as coisas são. Essa é uma questão difícil para o fundacionista cartesiano. Em particular, porque o critério de justificação para as crenças básicas, a infalibilidade, não é satisfeito por crenças sobre o mundo externo: há sempre a possibilidade de estarmos enganados sobre o que acreditamos acerca de objetos no mundo externo. Não é por menos, que essa versão de fundacionismo encontrada nos escritos de Descartes, também deu origem ao mais poderemos argumento cético sobre o mundo externo – o também chamado “ceticismo cartesiano”. Em suma, mesmo que for o caso que não podemos estar enganados sobre como as coisas nos parecem, não podemos derivar disso, com a mesma infalibilidade, algo sobre como as coisas são. Se isto estiver correto, então não temos conhecimento acerca das coisas do mundo externo.

Para outra versão do fundacionismo, a saber, o fundacionismo empírico, crenças básicas são aquelas que derivam sua justificação não de outras crenças, mas da experiência sensorial. Ou seja, crenças perceptivas, diretamente formadas com base na experiência sensorial, contariam todas como crenças

básicas. Por exemplo, a minha crença de que o café está amargo seria uma crença básica, para essa concepção de fundacionismo, porque é formada diretamente da sensação de amargo que experimento ao tomar o café. O leitor pode estar se perguntando se esta noção de crença básica não é a mesma apresentada anteriormente quando falamos de crenças de aparências. A resposta a essa questão, no entanto, é: não. Recorde que, para o fundacionismo cartesiano, crenças básicas são crenças que formamos sobre nossos estados mentais, que incluem crenças de aparência, ou seja, crenças sobre as nossas experiências sensoriais. Já para o fundacionismo empírico, crenças básicas são crenças formadas com base nas experiências sensoriais, mas não são crenças *sobre essas experiências*. Para tornar clara a diferença considere as seguintes crenças de S: (a) S acredita que o céu é azul; (b) S acredita que o céu lhe parece azul. A crença (a) é uma crença básica para o fundacionista empírico, uma vez que é uma crença formada com base na experiência de azul que S tem ao ver o céu. Mas (a) não é uma crença básica para o fundacionista cartesiano, uma vez que é uma crença sobre o mundo externo a S. Ou seja (a) é uma crença perceptiva e não uma crença introspectiva, i.e., uma crença sobre um estado mental de S. A crença (b), por outro lado, seria uma crença básica para o fundacionismo cartesiano por se tratar de uma crença de S sobre um estado mental, a saber, de S estar experienciando a cor azul ao olhar para o céu.

Tendo elucidado a diferença entre o fundacionismo empírico e o fundacionismo cartesiano, passemos para uma importante questão que o primeiro tem de responder. Como a experiência sensorial pode justificar uma crença? Para compreender a relevância dessa questão, considere que a noção de crença básica envolve duas características. Primeiro, uma crença básica é uma crença cuja justificação não deriva de outras crenças. Até aqui, não há problema, pois o fundacionista empírico alega que crenças perceptivas não dependem de outras crenças, para serem justificadas. Porém a segunda característica de uma crença básica é que ela é uma *crença justificada*. O fundacionista cartesiano explica essa segunda característica, sugerindo que crenças básicas possuem uma propriedade autojustificante (a infalibilidade, como vimos é a melhor candidata a essa propriedade), mas o fundacionista empírico toma um caminho diferente defendendo que a justificação de uma crença básica deriva da experiência sensorial. Assim, para o fundacionismo empírico ter uma resposta satisfatória para a questão “o que é uma crença

básica?”, é preciso explicar por que a experiência sensorial é uma fonte de justificação para uma crença. Sellars (1997) apresenta uma dura crítica a ideia de que a experiência sensorial pode contar como justificação: “ao caracterizar um episódio ou estado como o de conhecer (...) estamos colocando-o no espaço lógico das razões, da justificação e de ser capaz de justificar o que se diz”. O ponto de Sellars é que a justificação possui um caráter conversacional: se S está justificado em acreditar que p , então S deve ser capaz de apresentar as razões para sua crença em p . Mas a experiência sensorial pura não pode ser apresentada como uma razão para acreditar em algo, uma vez que se trata de uma experiência privada e não-conceitualizada. Em outras palavras, experiências sensoriais são incomunicáveis, portanto, não podem ser apresentadas como uma razão para se acreditar em algo. Uma forma de contornar o problema colocado por Sellars é a chamada “solução de compromisso” (STEUP, 2004): a ideia é a de que a experiência sensorial seja uma fonte de justificação para uma crença básica quando acompanhada de alguma razão que indique que essa experiência seja confiável. Por exemplo, tenho uma razão para acreditar que minha experiência perceptível é confiável para me mostrar objetos próximos e suas propriedades, porque, no passado, esse tipo de experiência me levou a formar crenças verdadeiras sobre objetos próximos. A razão que tenho para confiar em minhas experiências perceptivas pode ser usada no “espaço lógico das razões”, para apresentar a justificação para minhas crenças. Isso evita a objeção de Sellars, mas descaracteriza a motivação inicial do fundacionismo, isto é, a ideia de que há crenças básicas, que são justificadas sem recurso a outras crenças. Uma vez que a justificação para crenças perceptivas dependeria, além da experiência sensorial, de uma crença de que essas experiências são confiáveis, a solução de compromisso não é uma alternativa consistente com o fundacionismo, para evitar a objeção de Sellars (de fato, a solução de compromisso tem esse nome, por se tratar de um compromisso entre o fundacionismo e a sua tese rival, o coerentismo)

Mas se o problema com as experiências sensoriais for apenas a questão de não serem passíveis de serem apresentadas como razões que justifiquem uma crença, porque então não podemos conceitualizar ou descrever essas experiências e, assim, as tornar comunicáveis? Se optarmos por essa alternativa, retornamos ao fundacionismo cartesiano: ao descrever minha experiência de azul, ao ver o céu, eu diria coisas como “o céu me parece azul”, o

que torna nossa posição muito próxima daquela sobre as crenças de aparência, descritas anteriormente. Além de nos colocar face aos mesmos problemas enfrentados pelo fundacionismo cartesiano, Sellars também apresenta uma crítica adicional contra essa estratégia:

Quando digo que “X me parece verde” (...), o fato de eu fazer este relato, e não simplesmente “X é verde”, indica que certas considerações concorreram para suscitar, digamos, numa instância superior, a questão de “aceitar ou não aceitar”. Talvez tenha razões para pensar que X, afinal, pode não ser verde. (1997, p. 41)

A ideia de Sellars é a de que expressões do tipo “X me parece F” indicam incerteza ou dúvida quanto ao objeto X ter a propriedade F, e não indubitabilidade ou infalibilidade quanto ao conteúdo da crença. Portanto expressões desse tipo não podem ser tratadas como razões para se acreditar em algo. Assim, Sellars coloca o fundacionismo frente a um dilema (de fato, a objeção de Sellars ao fundacionismo é conhecida como “o dilema de Sellars”): experiências sensoriais não podem ser usadas como justificção para uma crença, uma vez que justificção tem a ver com a capacidade de fornecer razões para se acreditar, e experiências privadas e não conceitualizadas não são comunicáveis. Por outro lado, se descrevermos essas experiências, a fim de conceitualizá-las e torná-las comunicáveis, para serem usadas como razões para crer, o tipo de descrições que teremos não seriam boas razões, uma vez que seriam expressões de incerteza e dúvida.

A segunda estratégia evidencialista para solucionar Trilema de Agripa é o coerentismo. Para o coerentismo não existe algo como uma crença básica, ou seja, toda crença depende de outras crenças para ser justificada. A justificção de uma crença depende de ela manter uma relação de coerência com o sistema de crenças de um sujeito. Assim, para o coerentismo, uma crença C de S é justificada se C é coerente com o sistema de crenças de S. O que significa dizer que uma crença C mantém uma relação de coerência com um sistema de crenças? Para respondermos a isso, considere o seguinte exemplo: Maria é uma pessoa muito cientificista, ela acredita que apenas teorias comprovadas cientificamente são verdadeiras. Maria está passando por uma fase bastante

difícil em sua vida: ela perdeu o emprego e está endividada. Apesar de seu cientificismo, Maria costuma consultar o horóscopo e, apesar de acreditar que a astrologia não é comprovada cientificamente, ela dá crédito ao que o horóscopo diz: a má fase em sua vida se deve à influência da atual posição dos astros sobre pessoas de seu signo. Certamente, o leitor deve ter notado que há algo de errado com o sistema de crenças de Maria. De fato, as três crenças apresentadas são inconsistentes, ou seja, não podem ser todas verdadeiras. Se é verdade que apenas teorias comprovadas cientificamente são verdadeiras e que a astrologia não é comprovada cientificamente, então a astrologia não é uma teoria verdadeira sobre porque certas coisas ocorrem na vida das pessoas. Assim, em um primeiro sentido, podemos caracterizar a noção de sistema coerente de crenças quando não há inconsistência entre as crenças do sistema. Mas esse sentido fraco de inconsistência não é suficiente para uma boa caracterização de coerentismo sobre a justificação epistêmica. Isto porque qualquer teoria da justificação deveria rejeitar crenças inconsistentes mantidas por um sujeito como justificadas. Além disso, podemos ter um conjunto consistente de crenças, sem que este seja um conjunto coerente de crenças. Por exemplo, Maria acredita que o Céu é azul, que hoje é domingo e que o ventilador da sala está ligado. Essas três crenças são consistentes entre si, uma vez que todas podem ser verdadeiras, mas não parece o caso de formarem um conjunto coerente de crenças. Assim, além ausência de inconsistência, o coerentismo afirma que as crenças em um sistema coerente de crenças devem exibir uma relação de suporte mútuo. Por exemplo, as crenças de Maria de que hoje é domingo, de que a cafeteria da esquina está fechada e de que o comércio da cidade não abre aos domingos, seriam um conjunto coerente de crenças, uma vez que a verdade de uma dessas crenças confere suporte à verdade das outras.

O principal problema enfrentado pelo coerentismo é o chamado “problema do isolamento”. Um sistema de crenças pode ser coerente e não ter nenhuma relação com a realidade. A experiência não desempenha nenhum papel na justificação entendida nos termos do coerentismo, uma vez que alguém pode ter um conjunto coerente de crenças que não correspondam de forma alguma com a realidade à sua volta. A justificação epistêmica tem a ver com a ideia de que ela nos fornece alguma razão para pensar que nossas crenças são verdadeiras, mas a coerência de um conjunto de crenças parece não desempenhar esse papel. Bonjour (1985) oferece uma resposta ao problema do

isolamento, alegando que uma boa parte de nossas crenças são espontâneas, isto é, crenças formadas de maneira direta. A espontaneidade é característica de crenças perceptuais: ao olhar pela janela de meu apartamento eu formo uma série de crença de forma direta: acredito que já é noite, que alguns apartamentos vizinhos estão com as luzes acesas, que uma tempestade se aproxima etc. Essas crenças perceptuais mantêm um alto grau de coerência entre si e com meu sistema de crenças. Para Bonjour, isso nos dá uma boa razão para acreditar que elas são verdadeiras: uma vez que são crenças espontâneas, se fossem crenças desconexas com a realidade, elas poderiam representar o mundo de qualquer maneira, incluindo as maneiras mais incoerentes. Assim, a melhor explicação para a coerência de minhas crenças perceptuais é a de que são crenças verdadeiras.

2.3. Fontes individuais de justificação

A discussão sobre a justificação epistêmica na seção nos mostra que as fontes de justificação, quase que exclusivamente estudadas pela epistemologia tradicional, são aquelas que dizem respeito às capacidades cognitivas individuais dos sujeitos. Vimos que, para o fundacionismo cartesiano, uma importante fonte de justificação é a introspecção, que nos fornece crenças acerca de nossos estados mentais. Já para o fundacionismo empírico e para o coerentismo (para que o problema do isolamento seja evitado), a percepção é uma fonte de justificação relevante. Outras fontes de justificação individuais consideradas pela epistemologia tradicional incluem a intuição racional, para crenças sobre proposições matemáticas (como $2+2=4$), verdades lógicas (como $a=a$) ou verdades conceituais (todo solteiro é um não casado); a memória, para crenças formadas no passado (como, ontem faltou energia elétrica, passei por uma cirurgia em 2015, quando adolescente tinha um gato de estimação chamado Billy); e o raciocínio, para crenças que formamos a partir de inferências sobre crenças em outras proposições que julgamos verdadeiras (como a crença de que todas as esmeraldas são verdes, inferida da crença de que as esmeraldas observadas até agora são verdes). Cada uma dessas fontes individuais levanta uma série de questões epistemológicas interessantes, mas que estão fora do escopo deste trabalho. Uma sexta fonte de justificação, no entanto, que foi bastante negligenciada pela epistemologia tradicional, será de nosso interesse: o

testemunho. Nós rotineiramente formamos crenças com base naquilo que nos é dito por outras pessoas, sejam esses relatos escritos ou falados. Por exemplo, eu acredito que a temperatura irá cair nos próximos dias, em muitas regiões do país, porque alguém disse isso no noticiário da TV. Em epistemologia, chama-se “testemunho” qualquer relato que vise transmitir a informação de que p (como o relato do meteorologista sobre a previsão do tempo no noticiário) ou um relato que o ouvinte, ou o leitor, tome como informação de que p , ainda que este relato não tenha sido feito com a intenção de transmitir a informação – como um relato feito em um diário pessoal, utilizado posteriormente por um pesquisador, para escrever uma biografia sobre o autor do diário. Diferente das fontes de justificação mencionadas anteriormente, o testemunho é uma fonte social de justificação: testemunhos sempre pressupõem alguma comunicação entre dois ou mais indivíduos. A desconsideração do testemunho como fonte de justificação, na literatura epistemológica tradicional, é um dos fatores que revela o caráter individualista desta. Na seção seguinte eu pretendo especificar duas maneiras em que podemos entender o individualismo epistêmico e, posteriormente, mostrar como os trabalhos em epistemologia social questionam este pressuposto.

3. Individualismo Epistêmico

Há um sentido óbvio no qual o conhecimento humano possui um elemento social: muito do que sabemos, plausivelmente a maior parte do que sabemos, foi adquirido a partir do que outras pessoas nos ensinaram. Essa interação social mínima, a transmissão de informação de uma pessoa a outra, já poderia ser considerado um elemento social do conhecimento, o que legitimaria o campo de estudo da epistemologia social. Mas a epistemologia tradicional tem forte caráter individualista. O individualismo epistêmico pode ser caracterizado como a ideia de que o sujeito epistêmico deve empregar as suas próprias capacidades cognitivas, a fim de obter justificação para suas crenças e, assim, cumprir as condições para o conhecimento. Tal caráter individualista é mais pressuposto do que explicitamente defendido por epistemólogos tradicionais. Porém o individualismo pode ser encontrado de forma mais explícita em alguns clássicos da epistemologia. Por exemplo, John Locke, em uma passagem comumente citada para ilustrar o individualismo epistêmico, afirma que:

(...) opiniões de outrem em nossas mentes não nos fazem saber nem um pouco a mais, mesmo que verdadeiras. O que neles era ciência, em nós não passa de opinião, porque apenas demos assentimento à sua autoridade e não empregamos, como eles fizeram, nossa própria razão para compreender essas verdades (1975, i.iv, p.23).

O individualista epistêmico, portanto, parece assumir que crenças formadas em segunda mão, isto é, com base na palavra de outras pessoas, não constituem conhecimento, pois carecem de justificação. As crenças verdadeiras de um sujeito só estão justificadas se este empregar suas próprias faculdades cognitivas, a fim de avaliar a veracidade delas (essa interpretação individualista da passagem de Locke é sugerida por WELLBOURNE, 1981, para uma interpretação alternativa ver SHIEBER, 2009). Assim, podemos formular uma primeira versão do individualismo epistêmico pressuposto pela epistemologia tradicional da seguinte forma:

Individualismo Epistêmico₁: para toda proposição p conhecida por S, só é o caso que S sabe que p , se a justificação de S para crer que p for proveniente de uma fonte de justificação individual (i.e., intuição racional, introspecção, memória, percepção, raciocínio) de S.

Se o leitor se atentar, notará que o individualismo epistêmico₁ parece demasiado exigente. De fato, não parece sequer compatível com a maneira pela qual nós tipicamente aprendemos sobre a realidade à nossa volta. Nós supomos conhecer coisas sobre o passado distante, regiões do planeta que nunca visitamos, fatos que ocorreram em lugares em que não estávamos, episódios sobre nossa própria vida dos quais nos é impossível lembrar – como o dia em que nascemos ou o nome que recebemos ao sermos batizados. Também supomos ter algum conhecimento científico (ainda que elementar), como a composição química da água, sem que a maioria de nós estivesse em um laboratório de química, para empregar nossas próprias capacidades cognitivas, a fim de avaliar a veracidade da proposição “a água é H₂O”. O que há de comum nestes supostos casos de conhecimento é que não são casos onde a justificação

que os sustenta tem como fonte alguma das fontes de justificação individuais requeridas pelo individualismo epistêmico¹. Basta refletirmos sobre as muitas coisas que julgamos saber, para concluirmos que muito desse conhecimento teve como fonte – não a intuição racional, a introspecção, a memória, a percepção ou raciocínio – mas a *palavra de outra pessoa*, ou, como é tecnicamente chamado em epistemologia, o “testemunho” de outra pessoa. O testemunho, assim entendido, é uma fonte social de justificação epistêmica – o que, em princípio, a epistemologia tradicional de caráter individualista nega como fonte legítima. Mas se levarmos o individualismo epistêmico¹ a sério e rejeitarmos o testemunho como fonte de justificação, seríamos forçados a admitir que conhecemos muito pouco. Ou, colocando em outros termos, seríamos confrontados com o ceticismo sobre crenças formadas com base em testemunhos.

Uma segunda maneira de formular o pressuposto individualismo epistêmico da epistemologia tradicional é atentar para que tipos de sujeito a epistemologia trata como sujeitos epistêmicos. Por “sujeito epistêmico” eu me refiro a entidades capazes de instanciar atitudes doxásticas (como crenças) que possam ser avaliadas como justificadas, ou constituírem casos de conhecimento. Para a epistemologia tradicional o sujeito epistêmico sempre é um indivíduo. No entanto, as atribuições de conhecimento que empregamos na linguagem cotidiana não é consistente com este pressuposto. É comum atribuímos atitudes doxásticas e conhecimento a entidades coletivas, como grupos e instituições. Eis alguns exemplos encontrados em noticiários: os novos relatórios divulgados hoje destacam que a NASA **sabia** que, em quatro voos espaciais, desde 1983, o isolante se tinha soltado na parte dos tanques de combustível; uma equipe de cientistas **acredita** ter detectado, entre os dados enviados pelo observatório espacial Kepler, um planeta localizado a 2,8 mil anos-luz da Terra (...); (...) o governo brasileiro **sabia** oficialmente, 30 anos atrás, da morte do desaparecido político Ruy Carlos Vieira Berbert. Essas atribuições sugerem que coletividades também podem ser sujeitos epistêmicos. Mas esse tipo de conhecimento coletivo, como é chamado na literatura, não é levado em conta na epistemologia tradicional individualista. Atribuições de conhecimento coletivo, poderiam ser acomodados no enquadramento da epistemologia tradicional, como formas de se referir aos indivíduos que compõem o grupo. Por exemplo, como sugere Anthony Quinton (1976):

Diz-se que um grupo tem crenças, emoções e atitudes, que faz decisões e promessas. Mas estes modos de falar são claramente metafóricos. A atribuição de predicados mentais a um grupo é sempre um modo indireto de atribuir predicados a seus membros. Tais atribuições de estados mentais, como crenças e atitudes, são atribuições daquilo que chamo de um tipo de somatório. Dizer que a classe dos trabalhadores industriais está determinada a resistir às leis antisíndicas é o mesmo que dizer que todos ou a maioria dos trabalhadores industriais estão dispostos a isso (p. 02).

Quando atribuímos conhecimento de que p a algum grupo ou coletividade, o que queremos dizer, de fato, é que a maioria (ou todos) daquele grupo sabe (ou sabem) que p . Desta forma, em uma atribuição de conhecimento coletivo, o sujeito epistêmico ainda é o indivíduo e , portanto, não há algo como conhecimento coletivo. Essa ideia pode nos dar uma segunda formulação para o individualismo epistêmico pressuposto pela epistemologia tradicional:

Individualismo Epistêmico₂: apenas indivíduos podem ser sujeitos epistêmicos

Em grande medida, a epistemologia social surge como um desafio tanto ao individualismo epistêmico₁ quanto ao individualismo epistêmico₂. Isto não significa que um epistemólogo social precisa discordar de ambos os pressupostos. É possível admitir que alguns casos de conhecimento são baseados em fontes sociais, ou seja, rejeitar o individualismo epistêmico₁ e negar que entidades coletivas podem ser sujeitos epistêmicos, aceitando assim o individualismo epistêmico₂. De fato, as críticas às duas versões de individualismo epistêmico surgiram de forma independente uma da outra na literatura filosófica. Nas próximas seções, eu pretendo apresentar essas críticas como uma forma de ilustrar dois problemas centrais da epistemologia social, a saber, o problema da justificação por testemunho e o problema do conhecimento coletivo.

4. Epistemologia do Testemunho

Apesar do caráter individualista, a epistemologia tradicional não ignora completamente o fato de que muito do que sabemos tem como fonte o relato de outras pessoas. Por exemplo, David Hume, na Seção X das *Investigações Acerca do Entendimento Humano*, afirma: “[n]ão há raciocínio mais comum, mais útil e mesmo mais necessário à vida humana, que o derivado do depoimento humano, dos relatos das testemunhas oculares e dos espectadores”, (HUME, 1972, p. 102). O problema para a epistemologia tradicional de caráter individualista é explicar como esse tipo de conhecimento é possível, sem renunciar ao individualismo epistemológico. A resposta de Hume para esse problema é a de que a justificação por testemunho se funda em nossa experiência passada, que nos dá uma razão para acreditar em testemunhos. Por exemplo, se em minha experiência passada descobri, digamos, que taxistas geralmente dizem a verdade sobre endereços na cidade onde operam, então eu tenho uma razão para acreditar em taxistas, quando informam sobre endereços nesta cidade. Assim, a justificação que tenho para crer em relatos de taxistas ainda é proveniente de fontes de justificações individuais, a saber as minhas observações passadas, que confirmaram muitas vezes que relatos de taxistas sobre endereços são verídicos, e são inferências com base nessas observações que justificam a crença de que taxistas geralmente são confiáveis para informar sobre endereços.

Em contraste com David Hume, temos o pensamento de Thomas Reid, filósofo contemporâneo de Hume e um opositor do individualismo epistêmico. Nos *Essays on the Intellectual Powers of Man* (2011), Reid distingue operações mentais individuais das operações mentais sociais. As últimas diferem das primeiras, porque elas necessariamente pressupõem a comunicação entre dois indivíduos. Operações mentais sociais são atos como prometer, fazer um acordo, dar ou receber um testemunho. Para Reid, nossa confiança em testemunhos não deriva da experiência. Pois se assim fosse, só passaríamos a dar crédito nos relatos de outras pessoas, após termos aprendido pela experiência que as pessoas geralmente dizem a verdade. Mas o que parece ocorrer é o inverso: somos mais crédulos quando jovens e, à medida que ganhamos experiência, aprendemos em quais situações devemos ou não acreditar em alguém. Reid defende que temos uma propensão inata a acreditar naquilo que nos é dito, de

maneira análoga à propensão inata que temos em acreditar naquilo que percebemos pelos sentidos. Além disso, também temos uma propensão inata a sermos sinceros, ou seja, a dizer aquilo que acreditamos ser verdade. Se as pessoas tendem a dizer a verdade, então o testemunho de alguém de que p pode ser tomado como uma justificação para acreditar em p . Para epistemologia do testemunho de Reid, portanto, podemos presumir que uma pessoa é confiável, ao testemunhar, exceto nos casos em que houver alguma razão que indique o contrário.

O esboço anterior das posições de Hume e Reid correspondem às teses contemporaneamente chamadas, respectivamente, de “reducionismo” e “anti-reducionismo” em epistemologia do testemunho. Reducionistas afirmam que um ouvinte O está justificado em acreditar em uma testemunha T quando O tem razões para pensar que testemunhos são geralmente confiáveis. Antirreducionistas, por sua vez, defendem que O está justificado em acreditar em T quando não há razões contrárias para fazê-lo. Para reducionistas o testemunho funciona como um telescópio: temos justificação em acreditar em observações através de telescópios porque temos razões para acreditar que esse é um instrumento confiável para observações de longo alcance. Para antirreducionistas o testemunho é análogo à percepção: se observamos uma garrafa de água mineral cheia na geladeira, estamos justificados em acreditar que há água mineral na garrafa, mas se percebemos que o lacre da tampa da garrafa está rompido, temos uma razão contrária para acreditar que a água na garrafa é água mineral, talvez seja água da torneira. Neste caso, a percepção da garrafa apenas não fornece justificação para nossa crença sobre o seu conteúdo. Da mesma forma, se o dono da casa me informa que há água mineral na geladeira eu estou justificado, pela palavra dessa pessoa, a crer nisso. Se, por outro lado, o informante for um amigo, conhecido por mentir compulsivamente, eu não ganho qualquer justificação para acreditar. O reducionismo em epistemologia do testemunho é compatível com o individualismo epistêmico¹, uma vez que explica a justificação por testemunho, por redução a fontes individuais de justificação epistêmica. O anti-reducionismo, por outro lado, é uma posição que fornece suporte a reivindicação da epistemologia social, a saber, a de que há fontes sociais de justificação. Assim, uma defesa do anti-reducionismo em epistemologia do testemunho consiste, ao mesmo tempo, em uma defesa da epistemologia social. Nas próximas seções veremos as razões a

favor e contrárias a cada uma das teses.

4. 1. O embate reducionismo/anti-reducionismo em Epistemologia do Testemunho

Para compreendermos o embate entre reducionistas e antirreducionistas precisamos, primeiro, de uma ideia clara do que está em discussão. O desacordo entre as duas posições diz respeito a duas teses. A primeira tese diz que a mera declaração de que p proferida por T não fornece qualquer justificação para O acreditar em p . Vamos chamar essa tese de “Tese da Necessidade da Redução (R_N)”. Como vimos, reducionistas defendem R_N enquanto antirreducionistas rejeitam-na. No entanto, o reducionismo não pode ser identificado apenas com R_N . Isso porque reducionistas aceitam que frequentemente adquirimos conhecimento por testemunho. Por isso reducionistas também sustentam a tese de que é frequentemente possível para um ouvinte O obter razões para acreditar que testemunhos (ou certos tipos de testemunho) são geralmente confiáveis. Chamaremos a essa segunda tese de “Teste da Possibilidade da Redução (R_P)”. Desta forma, o reducionismo pode ser entendido como a conjunção de R_N e R_P . A posição do antirreducionista quanto a R_P , por outro lado, não é tão clara. Alguns antirreducionistas, como Thomas Reid, admitem que a experiência pode aumentar o grau de justificação fornecida por testemunho que se mostrou confiável no passado, mas a ausência de experiência passada que assegure a confiabilidade do testemunho não anula sua capacidade de fornecer justificação. Ou seja, a posição anti-reducionista é compatível com R_P : tudo que o anti-reducionismo diz é que o relato de T de que p fornece justificação para O acreditar em p (desde que não haja razões para O acreditar que T não é confiável), o que é compatível com a possibilidade de O ter razões para acreditar que T é confiável. Apesar disso, alguns epistemólogos críticos do reducionismo argumentam que a redução proposta pela R_P (pelo menos em um sentido particular, como veremos a seguir) não se sustenta. Portanto o anti-reducionismo pode se configurar tanto com a mera rejeição de R_N quanto com a rejeição de R_N e R_P . Outra combinação seria aceitar R_N e rejeitar R_P , o que nos conduziria ao ceticismo quanto ao conhecimento obtido por testemunho: se a declaração de uma testemunha de que p não fornecesse justificativa para acreditarmos em p e não é possível obter razões para cremos

justificadamente em T, então não é possível sabermos algo por testemunho. A tabela a seguir sumariza as posições, de acordo com a aceitação ou rejeição das teses da Necessidade da Redução e da Possibilidade da Redução.

	Necessidade da Redução	Possibilidade da Redução
<i>Reduccionismo</i>	Verdadeira	Verdadeira
<i>Anti-reduccionismo</i>	Falsa	Verdadeira
<i>Anti-reduccionismo</i>	Falsa	Falsa
<i>Ceticismo</i>	Verdadeira	Falsa

4.2. Argumentos acerca da Possibilidade da Redução

Podemos encontrar um tipo de defesa de R_P nas *Investigações Acerca do Entendimento Humano*, onde David Hume nos fornece uma explicação de quando o testemunho de que p pode ser considerado como prova ou probabilidade de que p :

Como o depoimento que deriva das testemunhas e do testemunho humano se funda sobre a experiência passada, varia com a experiência e se considera ou uma *prova* ou uma *probabilidade*, conforme se tem verificado constante ou variável a conjunção entre um gênero particular do relato e um gênero do objeto (HUME, 1972, pp. 102-103. Itálicos do original).

Hume parece sugerir que testemunhos podem ser classificados em gêneros ou tipos. Por exemplo, diagnósticos médicos, previsões do tempo em noticiários, história de pescadores... A experiência passada nos diz se um tipo de testemunho costuma ser ou não confiável. Se a experiência passada nos mostra que um tipo de testemunho costuma ser confiável, então estamos justificados em acreditar em testemunhos deste tipo. Dessa forma, o reduccionismo humeano sugere que a justificação por testemunho funciona da seguinte maneira:

1. T testemunha a O que p ;
2. A experiência passada mostrou a O que testemunhos do tipo T geralmente são confiáveis.
3. Então O está justificado por (2) em acreditar que testemunhos do tipo T geralmente são confiáveis.
4. Portanto, O está justificado por (3) em acreditar que p .

No ensaio “Testimony and Observation” (1973) e, posteriormente, no livro *Testimony: A Philosophical Study* (1992), C.A.J. Coady, oferece uma série de objeções contra o reducionismo humeano. Em linhas gerais, o argumento desenvolvido por Coady, explora a ambiguidade no termo “experiência” empregado por Hume em sua explicação da R_p . Conforme vimos, para Hume estamos justificados em acreditar em testemunhos de um determinado tipo, se a experiência passada mostrou que esse tipo de testemunho é geralmente confiável. No entanto, Coady aponta que “experiência”, nesse contexto, pode ter dois sentidos. Em um primeiro sentido, “experiência” significa a experiência comum, partilhada pelos seres humanos. Por exemplo, na afirmação “a experiência passada nos diz que um homem morto não pode voltar à vida”, a palavra “experiência” aqui pode se referir à experiência comum: nunca se ouviu falar que mortos voltaram a vida. Se for esse o sentido de “experiência” empregado por Hume, sua explicação da R_p incorre em circularidade, uma vez que a noção de experiência comum pressupõe o testemunho de outras pessoas. Seria como se eu dissesse que a razão pela qual acredito nos outros é porque me disseram que são confiáveis! Por outro lado, em um segundo sentido, “experiência” pode se referir à experiência individual: em *minha* experiência passada, *eu* nunca observei um homem morto voltando a vida. Porém, se Hume, com “experiência”, pretende se referir à experiência individual, sua abordagem da R_p não consegue explicar o vasto conhecimento que supomos ter por testemunho: confirmamos a veracidade de alguns tipos básicos de testemunho – confirmei que meus vizinhos são precisos para informar sobre onde consigo comprar os mais variados produtos domésticos nas redondezas – mas, para certos tipos de testemunho, a confirmação da veracidade parece não ser frequente. Testemunhos científicos, testemunhos sobre regiões que nunca visitamos, sobre eventos históricos, são exemplos de tipos de testemunhos que

em geral não confirmamos a veracidade. Se R_P for entendida em termos de experiência individual, então, em geral, não podemos acreditar justificadamente em muitos tipos de testemunho. Na literatura, chama-se a esse problema de “problema da pouca base empírica”, uma vez que diz respeito à pouca quantidade de confirmação empírica que temos para assegurar a confiabilidade de testemunhos em geral.

Outro problema também relacionado à pouca base empírica para assegurar a confiabilidade de testemunhos, que podemos chamar de “problema do testemunho em fase infantil”, é que crianças dificilmente ganhariam justificação por testemunho nesta interpretação da R_P . Por terem pouca experiência, é plausível pensar que crianças confirmaram a veracidade de um número ainda menor de testemunhos, mesmo dos tipos mais básicos. Uma vez que crianças, em princípio, aprendem muito sobre o mundo a sua volta através do testemunho de seus pais, familiares e professores, a abordagem humeana de R_P se torna ainda mais implausível. Note que o problema da pouca base empírica e o problema do testemunho em fase infantil sugerem que R_P , entendida nos termos humeanos, é falsa. Uma vez que R_P não diz apenas que é possível para O obter razões para acreditar que certos tipos de testemunhos geralmente são confiáveis, mas que é *frequentemente* possível para O fazer isso. Além disso, uma vez que o reducionismo admite R_N , as objeções de Coady a R_P , se funcionarem, nos conduzem ao ceticismo quanto ao conhecimento obtido por testemunho. Se quisermos evitar o ceticismo, sugere Coady, devemos também negar R_N , aceitando o testemunho como uma fonte legítima de justificação. Antes de investigarmos, no entanto, se temos boas razões para aceitarmos ou rejeitarmos R_N , veremos outra abordagem de R_P que, em princípio, evita as dificuldades colocadas por Coady.

Em uma série de trabalhos, Elizabeth Fricker (1987, 1994, 1995, 2002, 2004) defende uma interpretação menos exigente de R_P a qual ela denomina de “redução local” (em contraste com o que ela chama de “redução global”, para se referir ao reducionismo do tipo humeano que vimos anteriormente). A diferença entre o reducionismo local e o reducionismo humeano é que a justificação para acreditar em um testemunho não depende de uma crença de que testemunhos ou tipos de testemunhos são geralmente confiáveis, mas que uma testemunha em particular é confiável quanto a um relato em particular. Por se referir a razões para acreditar em um testemunho em particular, a confiabilidade de uma

testemunha pode ser justificada por outros testemunhos sem incorrer em circularidade (como ocorria no caso da redução humeana). Por exemplo, posso estar justificado em acreditar que um mecânico em particular é confiável para me informar sobre qual é o defeito no motor de meu carro, com base no testemunho de meu pai de que este mecânico, que trabalha para a família há anos, é confiável. Uma vez que o testemunho de outras pessoas podem ser usados para justificar a crença na confiabilidade de um testemunho em particular, em princípio, podemos dizer que essa versão de R_p contorna também o problema da pouca base empírica, enfrentado pelo reducionismo humeano, uma vez que a base para estabelecer que um testemunho em particular é confiável seria maior por incluir outros testemunhos.

Porém, o reducionismo local ainda enfrenta problemas semelhantes ao reducionismo global. Considere, por exemplo, que em muitos casos de testemunhos não temos conhecimento prévio algum sobre a testemunha em particular. Quando pergunto a um estranho acerca da localização de um hotel eu não tenho qualquer crença que indique que essa testemunha particular é confiável para esse tipo de relato. Ainda assim, parece plausível pensar que estou justificado em acreditar naquilo que o desconhecido informante de rua me diz sobre a localização do hotel. Mesmo nas situações onde a confiabilidade de alguém pode ser justificada com base no testemunho de outra pessoa, em algum momento na cadeia de justificação, o sujeito teria que tomar um testemunho como confiável sem ter outras razões que assegurem isso. O problema do testemunho em fase infantil também se mantém para esta versão de R_p , já que o escopo de crenças que assegurem a confiabilidade de instâncias particulares de testemunho seria muito pequeno para crianças e jovens. Fricker está ciente desses problemas em sua elaboração de R_p , mas a solução que ela apresenta é no mínimo curiosa. Em relação ao testemunho em fase infantil, Fricker afirma que “nós temos de aceitar (o anti-reducionismo) com respeito à fase de desenvolvimento, mas não precisamos e não devemos aceitar com respeito à fase madura” (1995, p. 403), e também o mesmo com respeito a demais situações, onde ouvintes não têm qualquer razão que assegure a confiabilidade da testemunha: “(a palavra de alguém) teria papel ativo (na justificação) somente em circunstâncias incomuns onde alguém não sabe nada exceto que outro está testemunhando que p ” (2002, p. 382). Fricker então parece admitir que, em casos de testemunho em fase infantil e em casos onde o ouvinte não tem

alguma razão para acreditar que a testemunha é confiável (que não seja o próprio relato da testemunha), a declaração de uma testemunha de que p fornece justificção para um ouvinte acreditar que p . Ou seja, Fricker parece admitir que R_N é falsa. Se esse for o caso, então a posição de Fricker não confere suporte a uma tese reducionista. Conforme visto, o reducionismo é a conjunção de R_N e R_P e, uma vez que Fricker parece rejeitar R_N em algumas situações não claro que sua posição seja de fato reducionista. Colocando em outros termos: se, para o reducionismo local, a palavra de alguém é o bastante para acreditarmos nela em algumas circunstâncias, então por que o reducionismo local é um reducionismo?

4.3. Argumentos acerca da Necessidade da Redução

Reducionistas defendem que a mera declaração de uma testemunha T de que p não fornece justificção para O acreditar que p . A Tese da Necessidade da Redução, como vimos, é o coração do reducionismo. Mas porque devemos aceitá-la como verdadeira? Por que a palavra de alguém de que p não basta como uma razão para acreditarmos em p ? Um argumento em favor de N_R diz respeito a uma diferença supostamente crucial entre o testemunho e outras fontes individuais de justificção como, por exemplo, a percepção. Se formamos uma crença falsa com base na percepção (ou em outra fonte individual de conhecimento), tipicamente a falsidade pode ser explicada por algum erro de interpretação (como quando o sujeito se confunde ao ver algo) ou mal funcionamento (como quando o sujeito tem uma alucinação). Por outro lado, uma crença formada a partir de um testemunho pode ser falsa mesmo quando não há qualquer erro de interpretação ou funcionamento: o ouvinte pode entender perfeitamente bem o que a testemunha intencionou dizer e ainda assim a crença formada pelo ouvinte pode ser falsa porque a testemunha deliberadamente disse uma falsidade. Em suma, a testemunha pode mentir. Essa diferença foi observada mesmo por Thomas Reid, que tipicamente é considerado um anti-reducionista, em sua abordagem do testemunho:

Homens às vezes nos conduzem a erros quando entendemos perfeitamente sua linguagem, por nos dizerem mentiras. Mas a natureza nunca nos

engana dessa forma; sua linguagem é sempre verdadeira, e somente quando é má interpretada é que nós caímos em erro (REID, 1764, p. 123).

Uma vez que T pode intencionalmente declarar uma falsidade, argumenta o reducionista, é necessário que O tenha razões para pensar que T é confiável a fim de estar justificado em acreditar no que T diz. Desta forma, temos uma razão em favor de N_R . Uma objeção a esse argumento, no entanto, é a de que a intenção de alguém de enganar pode afetar mesmo crenças formadas com base na percepção. Por exemplo, um assaltante pode usar uma réplica de uma arma de fogo para fazer suas vítimas acreditarem falsamente que ele está armado. Green (2006) argumenta que objetos falsos são análogos perceptivos a mentiras em casos de testemunho. O reducionista pode rejeitar essa objeção, argumentando que casos de objetos falsos são raros, enquanto mentiras são comuns. Neste sentido, Jennifer Lackey (2006) argumenta que “falhas no caso do testemunho são muito mais *imprevisíveis* que falhas em casos não-testemunhais” (p. 176, *italicos do original*). Afirmar que testemunhos são imprevisíveis, porém, tem um custo alto para o reducionista. Isto porque, conforme argumenta Peter Graham (2003), se testemunhos forem imprevisíveis isso implicaria na falsidade de R_P . Lembre-se que R_P afirma que é frequentemente possível para O obter razões para acreditar que testemunhos são confiáveis, mas isso pressupõe alguma previsibilidade com relação a confiabilidade de testemunhos. Para entender como esse é o caso, considere o seguinte exemplo. Suponha que alguém lance um dado comum dezenas de vezes e (por uma mera obra do acaso) obtém o valor 3 em todas as rolagens. Isso não confere justificção para alguém acreditar que o dado voltará a cair 3, se for rolado novamente. Afinal, se tratando de um dado comum, as chances de um valor 3 ser rolado continua sendo de uma em seis. De forma análoga, se testemunhos forem imprevisíveis, mesmo que a testemunha diga a verdade dezenas de vezes antes, isso não confere justificção para alguém acreditar nela na próxima vez que ela testemunhar algo. Portanto, se testemunhos são imprevisíveis, não é possível para O ter razões para pensar que testemunhos são confiáveis, mesmo em uma circunstância particular, o que tornaria R_P falsa, mesmo na versão do reducionismo local. Mas será mesmo que temos boas razões para pensar que testemunhos são imprevisíveis? De que as pessoas são

livres para mentir ou dizer a verdade, não se segue que não podemos prever se elas irão dizer a verdade ou não. O comportamento pode ser previsível, mesmo supondo que seres humanos são livres para agirem. Quando nos dirigimos ao supermercado, por exemplo, consideramos que os funcionários do supermercado, apesar de serem livres, se comportam de forma previsível: eles abriram o supermercado no horário usual e estão atendendo de maneira usual. Agimos no mundo com relação a outras pessoas, prevendo que estas se comportarão de certa maneira, mesmo julgando que elas são livres para agir de muitas outras maneiras. Assim, argumenta o anti-reducionista, podemos presumir que as pessoas dirão a verdade, mesmo sendo livres para mentir, quando não há razões que indiquem o contrário.

Outro argumento em favor de N_R é o argumento contra a credulidade (FRICKER, 1994). A estratégia reducionista aqui tem dois passos: primeiro argumenta-se que a credulidade é uma atitude epistemicamente viciosa e, assim, incompatível com a noção de justificação epistêmica. Depois, mostra-se que se aceitarmos o anti-reducionismo, isto é, se admitirmos que a declaração T de que p fornece justificação para O acreditar que p na ausência de razões contrárias, estaríamos sancionando a formação de crenças crédulas. Portanto, não pode ser o caso que a declaração T de que p fornece justificação para O acreditar que p (mesmo na ausência de razões contrárias), exatamente como afirma N_R . Vejamos o primeiro passo.

Há várias maneiras de formulamos a noção de credulidade como uma atitude epistemicamente viciosa. Para fins de argumentação, consideraremos apenas a credulidade quanto à crença em testemunhos (cf., GOLDBERG e HENDERSON, 2005). Podemos dizer que O é crédulo quanto a testemunhos se (a) O acredita em qualquer coisa que lhe dizem ou (b) O frequentemente ignora razões que indicam que a testemunha não é confiável. Porque a credulidade, assim entendida, seria um vício epistêmico? Porque alguém crédulo, neste sentido, plausivelmente formaria muitas crenças falsas com base em testemunho. Desta forma, a credulidade viola a meta epistêmica que, conforme vimos, é a de acreditar em verdades e evitar falsidades. Mas para que isso conte como uma razão em favor de N_R o reducionista deve mostrar que, se admitirmos que, na ausência de razões contrárias, a palavra de alguém de que p fornece justificação para O acreditar que p , estaríamos sancionando a credulidade. Para sustentar essa ideia Fricker afirma que “em uma abordagem reducionista, mas não em

uma abordagem anti-reducionista, o ouvinte deve sempre estar monitorando criticamente a testemunha” (1994, p.143). Assim, apenas o reducionismo evitaria a credulidade. No entanto, o que Fricker parece sugerir é que o anti-reducionismo é a tese de que temos justificação para confiar cegamente em testemunhos, mas tal interpretação não faz justiça aquilo que o anti-reducionista defende. Como aponta Coady:

(...) a recepção do testemunho é normalmente irreflexiva, mas não acrítica (...) nós podemos simplesmente reconhecer que sinais padrões de engano, confusão, ou erro não estão presentes. Esse reconhecimento incorpora nosso conhecimento da competência da testemunha, das circunstâncias que envolvem sua declaração, da sua honestidade, da consistência das partes de seu testemunho, e sua relação com o que outros disseram ou não disseram sobre o assunto (1992, p. 47).

O anti-reducionismo diz que a palavra de uma testemunha é fonte de justificação para um ouvinte apenas quando não há razões que indiquem ao ouvinte que a testemunha não é confiável. Se houver razões indicando a inconfiabilidade de T, então O não está justificado em acreditar na palavra de T mesmo se O ignorar (por credulidade) essas razões. Portanto o anti-reducionismo, ao contrário do que sugere Fricker, não é compatível com a credulidade.

Considerando a maneira pela qual aprendemos sobre o mundo através de testemunhos, as dificuldades em explicar esse conhecimento, por meio de fontes individuais de justificação, como sugerem R_P e as questionáveis razões para aceitarmos R_N , parece razoável admitir o anti-reducionismo sobre a justificação por testemunho. Se isto estiver correto, temos boas razões para rejeitar o individualismo epistêmico₁: para muitas proposições p que conhecemos, não é o caso que a justificação que temos para acreditar em p é proveniente de uma fonte de justificação individual. Resta agora nos perguntar se o individualismo epistêmico₂ está correto. Ou seja, vamos investigar se há entidades epistêmicas coletivas.

5. Crenças coletivas e a atribuição de conhecimento a grupos

Com vimos, a epistemologia tradicional de caráter individualista assume que sujeitos epistêmicos são sempre indivíduos. Nós também vimos que tal pressuposto é inconsistente com a nossa prática linguística de atribuir conhecimento a entidades coletivas. A epistemologia tradicional tem uma boa razão para não rejeitar esse pressuposto, a saber, a própria análise clássica de conhecimento. Na introdução deste texto, foi dito que a análise clássica estipula que conhecimento é, pelo menos, crença verdadeira justificada. O elemento problemático nas atribuições de conhecimento coletivo é a crença. Tipicamente, crenças são entendidas como estados mentais, o que nos leva a pensar que apenas sujeitos com cérebros e mentes podem ter crenças. Seria muito estranho considerarmos que grupos são um tipo de entidade que possui mentes, já que certamente grupos não tem cérebros. Por isso, não parece plausível considerar grupos como entidades que têm crença. Seria então o caso que nossas atribuições de conhecimento a grupos na linguagem cotidiana não são literais? Ou seria o caso que conhecimento não implica crença? Ou quem sabe grupos podem ter crenças? A epistemologia coletiva, como é conhecida a subárea da epistemologia social que investiga questões epistemológicas relacionada a grupos, surge com uma tentativa de resposta à última questão. Em particular nos trabalhos de Margaret Gilbert (1987, 1989, 1994, 2004), que argumenta persuasivamente em favor da ideia de que grupos podem ter crença. Nas seções seguintes, eu reconstruirei em partes o debate envolvendo a noção de crença coletiva.

5.1. Abordagem Somatória de Crenças Coletivas

O individualismo epistêmico pode, em princípio, argumentar que atribuições de conhecimento ou crenças coletivas são, em verdade, modos de se referir a soma do conhecimento ou crenças individuais dos membros que compõem um grupo, tal como foi observado na sugestão de Anthony Quinton (Seção 1). Nesta seção, investigaremos se tal sugestão se sustenta e observaremos algumas particularidades acerca do fenômeno ao qual chamamos de “crenças coletivas” que, acredito, serem persuasivas o bastante para mostrar que a noção de crença coletiva, apesar de soar metafisicamente estranha, não é

de todo absurda. Para iniciar a investigação, vamos tornar preciso aquilo que é expresso no pensamento de Quinton, que doravante, será chamado de “visão somatória de crenças coletivas”. Segundo a visão somatória uma crença coletiva é definida da seguinte maneira: um grupo G acredita que p se, e somente se, a maioria dos membros de G acredita que p . Posto dessa forma, a condição de que a maioria dos membros de G acredite que p é uma condição *necessária e suficiente* para que seja o caso de que G acredite que p . Como resultado, a crença coletiva é redutível à soma das crenças individuais. Assim, o segundo pressuposto do individualismo epistêmico se mantém: quando dizemos que um grupo acredita que p , queremos dizer, com isso, que a maioria dos membros daquele grupo tem a crença de que p . O sujeito epistêmico, portanto, ainda é o indivíduo. A visão somatória, no entanto, sofre com uma série de contraexemplos. Examinaremos alguns a seguir.

Um primeiro problema para a visão somatória é que a condição de que a maioria dos membros de G acredite que p não parece ser suficiente para que G acredite que p . Isso pode ser ilustrado com o seguinte contraexemplo: imagine um grupo de trabalhadores de uma montadora de carros. Suponha que a maioria desses trabalhadores acredite que seus salários e condições de trabalho estejam muito ruins e que, o melhor a se fazer para resolver essa situação é que os trabalhadores entrem em greve. Uma vez que a visão somatória de crenças coletivas nos diz que se a maioria dos membros de um grupo acredita que p , então o grupo acredita que p , segue-se que, segundo a visão somatória, que é correto afirmar que o grupo de trabalhadores acredita que o melhor a se fazer, para se resolver o problema das condições de trabalho e do baixo salário, é entrarem em greve. Mas acrescenta a este cenário a seguinte consideração: cada um dos trabalhadores que acredita que eles devem entrar em greve mantém essa crença em segredo, talvez por receio de que se falarem sobre isso serão demitidos. Segundo Gilbert (1987), se cada um dos membros do grupo mantém em privado a crença de que p , não pode ser o caso de que o grupo acredite que p . Mas por que não? Para alguns pensadores uma crença coletiva genuína deveria capaz, em princípio, de motivar uma ação coletiva do grupo. Assim como uma crença individual motiva uma ação individual. Anthonie Meijers (2002) argumenta que os membros de um grupo supostamente teriam de defender sua crença, quando questões relativas a ela forem discutidas, ou quando decisões com base nela forem tomadas. Membros de um grupo, escreve

Meijers “também teriam o direito de reivindicar que os outros membros façam o mesmo: sua crença coletiva cria comprometerimentos recíprocos” (2002, p. 73). Se os membros de um grupo mantêm em privado uma determinada crença, nenhum comprometimento recíproco surgiria, nenhuma ação ou atitude coletiva com relação a crença ocorreria. Em virtude disso, não parece ser o caso que o grupo de trabalhadores acredita que devem fazer greve, ainda que a maioria deles pessoalmente acredite nisso.

Mas um defensor da visão somatória poderia argumentar que atribuições de crenças a grupos são metafóricas e por isso não precisa explicar qualquer efeito de comprometimento ou atitude coletiva. Afinal, se atribuições de crenças coletivas são uma maneira de se referir as crenças individuais, não é necessário explicar qualquer fenômeno coletivo com base nelas. Apesar de essa ser uma saída para alguém que rejeite a existência de crenças coletivas genuínas, há outras considerações que temos de ter em mente, antes de nos rendermos a uma visão radicalmente metafórica. Tollefsen (2002) argumenta que temos pelo menos duas razões para pensar que atribuições de crenças coletivas não são meras ficções. Primeiro, considere a nossa prática de atribuir responsabilidade a grupos como organizações e corporações, como, por exemplo, a responsabilidade legal da Samarco pelo rompimento da barragem de Mariana. Atribuir responsabilidade a uma organização parece pressupor que essas têm atitudes intencionais e crenças (que são componentes das ações pelas quais a organização é responsabilizada). Dificilmente alguém pensaria que a Samarco é apenas metaforicamente responsável pelo desastre ambiental em Mariana. Isso sugere que atribuições de crenças a grupos são *prima facie* não-metafóricas. Segundo atribuições de crenças e outros estados intencionais a grupos têm grande poder explicativo. Imagine se, com uma modificação no exemplo anterior, um grupo de trabalhadores deflagrasse greve. Os trabalhadores tomariam uma série de ações coletivas, como parar as atividades, reivindicar reuniões com seus empregadores a fim de exporem suas exigências etc. Além disso, o grupo cobraria que cada trabalhador cooperasse com essas ações. A crença coletiva de que os trabalhadores devem fazer greve consegue explicar o comportamento coletivo membros do grupo. Evidentemente, isso não é o bastante para afirmar que essas atribuições de crenças a grupos são literais, mas fornece um indício para não tratarmos crenças coletivas como meras metáforas ou modos de dizer. Portanto precisamos distinguir entre uma visão de

que crenças coletivas são puramente metafóricas (que podemos chamar de “ficcionalismo”) e uma visão somatória que admite que há crenças coletivas, mas que essas são a soma das crenças individuais e outra condição que as torne crenças coletivas. O contraexemplo das crenças privadas mostra que a ideia de que crenças coletivas são *apenas* a soma das crenças individuais é falsa. Portanto o defensor da visão somatória precisa procurar por uma condição adicional que torne uma mera soma de crenças numa crença coletiva.

Uma forma de o defensor da visão somatória evitar o problema levantado pelo contraexemplo das crenças privadas é incluir uma condição que exija que os membros de um grupo tenham ciência da crença individual de cada um na análise da crença coletiva. Isto pode ser feito, lançando mão de uma concepção de *conhecimento comum*. David Lewis (1969) nos fornece uma definição que pode servir a esse propósito: há conhecimento comum de que p entre os membros de um grupo G quando (i) é o caso que p ; (ii) todos em G sabem que p ; (iii) todos em G sabem que (ii); e assim por diante. Desta forma, podemos reformular a análise inicial de crença coletiva, em uma visão somatória, adicionando a condição de conhecimento comum: Um grupo G acredita que p se, e somente se, (1) a maioria dos membros de G acredita que p ; (2) é conhecimento comum em G de que (1). Com a adição da condição de conhecimento comum, casos de crenças privadas seriam excluídos: só há uma crença coletiva de que p quando membros do grupo em questão sabem que cada membro acredita que p . Desta forma, retornando ao nosso contraexemplo anterior, os trabalhadores não acreditam *coletivamente* que devem entrar em greve, porque não cumprem a condição (2). Em princípio, isso soluciona o problema; porém há outros contraexemplos que dificultam a posição do defensor da visão somatória, mesmo nessa reformulação. Consideremos o caso de grupos coextensivos apresentado por Gilbert (1987)

Imagine duas comissões – a Comissão dos Alimentos e a Comissão da Biblioteca de uma Faculdade local – com os mesmos membros. Parece possível dizer, sem contradição, que (a) a maioria dos membros da Comissão da Biblioteca pessoalmente acredita que os alunos da Faculdade têm que consumir mais amido. Isto é conhecimento comum entre os membros desta

comissão; (b) acontece exatamente o mesmo, *mutatis mutandis*, com os Membros da Comissão dos Alimentos; (c) a Comissão dos Alimentos acredita que os alunos da Faculdade têm que consumir mais amido, por outro lado, a Comissão da Biblioteca não tem opinião sobre o assunto (p. 187).

Com esse exemplo, Gilbert pretende demonstrar que, supondo dois grupos, G1 e G2, formados pelos mesmos membros, é possível que a maioria dos membros de G1 e G2 acredite que p , que seja conhecimento comum em entre os membros de G1 e G2 e, no entanto, pode ser o caso que G1 acredite que p e não seja o caso que G2 acredite que p . Isso se configura em um contraexemplo à versão de conhecimento comum da visão somatória de crenças coletivas, porque G2 cumpre as duas condições estipuladas por essa versão, mas isso não é suficiente para que G2 tenha uma crença coletiva de que p . Essa versão de crença coletiva somatória também não consegue explicar o efeito de comprometimento que crenças coletivas parecem exercer. Além disso, a noção de grupo (como um sujeito coletivo), não parece ter qualquer papel relevante nesta formulação de crença coletiva. Por exemplo, é plausível que a maioria das pessoas maiores de 18 anos acredita que se colocar a mão no fogo irá se queimar, assim como também é plausível que isto seja conhecimento comum entre as pessoas maiores de 18 anos. Mas não parece ser o caso que haja aqui um grupo que acredita que se colocar a mão no fogo irá se queimar: não há nada que motive a união e o comprometimento para ações e atitudes coletivas relativas a essa crença. John Searle (2011) ilustra este problema com o seguinte exemplo:

Um grupo de graduados da Harvard Business School que não só aprendeu, mas tornou-se partidário da teoria da mão invisível de Adam Smith. Após o dia da formatura, todos saíram ao redor do mundo e tentaram beneficiar a humanidade agindo da maneira mais egoísta possível e tentando tornarem-se os mais ricos que pudessem. Cada um tem conhecimento mútuo de que o outro está fazendo a sua parte. Assim, cada

um tem a mesma meta e cada um sabe que todos os outros sabem que tem esta meta e cada um sabe que todos sabem que cada um sabe disso. Porém, não existe cooperação. (...) neste caso as pessoas têm uma meta e estas pessoas têm um conhecimento comum de que as outras pessoas têm essa meta, mas, a meu ver, não há intencionalidade coletiva (p. 47–8)

Portanto, a visão somatória, mesmo com a adição da condição de conhecimento comum, parece sofrer do mesmo problema da primeira versão que analisamos. Podemos concluir, a partir dessas considerações, que a visão somatória não é suficiente para uma análise de crença coletiva. Poderíamos, é claro, procurar por mais condições, para acrescentar à definição, a fim de torná-la suficiente. Mas não seguirei por esse caminho. Ao invés disso, apresentarei um contraexemplo ao pilar que sustenta a visão somatória, a saber, a ideia de que uma condição necessária para o conceito de crença coletiva é a de que a maioria dos membros de G acredite que p , a fim de que seja o caso que G acredite que p . Se o contraexemplo for bem-sucedido, poderemos concluir que é possível que um grupo acredite que p , sem que a maioria de seus membros (talvez nenhum de seus membros!) acredite pessoalmente que p . Desta forma, há um tipo de crença que é essencialmente social, uma vez que é propriedade de um grupo e não de um indivíduo – o que refutaria o segundo pressuposto do individualismo epistêmico.

Considere um comitê composto por três membros formado para a contratação de um novo funcionário na empresa que gerenciam. O comitê seleciona dois candidatos, João e Maria, a partir de currículos enviados à empresa, e os submete a uma entrevista e a uma prova. Considerando a entrevista que os membros do comitê realizaram com os candidatos, cada um deles acredita que João é o melhor candidato ao cargo. Mas, considerando apenas a pontuação da prova e a análise do currículo, Maria é a melhor candidata. O comitê então decide excluir a pontuação da entrevista e levar em conta apenas a pontuação da prova e a análise do currículo na seleção do melhor candidato, por considerarem a entrevista um método epistemicamente falho (cf. MATHIESEN, 2011). A despeito disso, os membros do comitê não conseguem deixar de acreditar que João é o melhor candidato, com base na

forte impressão que ele causou na entrevista. Porém, uma vez que a decisão do comitê irá se basear apenas na pontuação da prova e análise do currículo, Maria é selecionada. Nesta situação, expressões como “O comitê *decidiu* que Maria é a melhor candidata a vaga de emprego” ou “Este comitê *acredita* que Maria é a melhor candidata para a vaga de emprego” soam perfeitamente apropriadas ainda que nenhum dos membros do comitê acredite *pessoalmente* nisso. Mais do que isso, neste caso, o posicionamento do comitê assegura ações coletivas apropriadas, como a contratação de Maria. Se isso estiver correto, temos um caso onde um grupo G acredita que p e nenhum dos membros de G acredita que p . O que mostra que a condição de que a maioria dos membros de G deve acreditar que p para que G acredite que p não é uma condição necessária para a análise da crença coletiva.

5.2. Abordagem de comprometimento conjunto de Crenças Coletivas

O que é então uma crença coletiva? Gilbert oferece uma resposta a essa questão através da noção de comprometimento conjunto. Para Gilbert um Grupo G acredita que p se, e somente se, membros de G estão conjuntamente comprometidos em acreditar que p como um corpo (GILBERT, 2004, p. 100). A noção de “comprometimento” deve ser entendida no seguinte sentido: suponha que alguém decida frequentar à academia três vezes na semana. Isto significa que esta pessoa está comprometida a ir à academia três vezes na semana. Neste caso temos um comprometimento pessoal, o que implica que o comprometimento pode ser rescindido pela decisão da própria pessoa que o fez. Um comprometimento conjunto, por outro lado, envolve dois ou mais indivíduos. Porém, não deve ser entendido como a mera soma ou a conjunção de comprometimentos individuais. Digamos que eu decida frequentar a academia três vezes por semana e meu vizinho decida o mesmo. Ambos estamos comprometidos com ir à academia três vezes por semana, mas este não é um caso de comprometimento conjunto. Se meu vizinho rescindir o comprometimento, mudando a frequência ou mesmo deixando de ir à academia, isto não me confere o direito de reclamar que o comprometimento foi violado. Nas palavras de Gilbert, “um comprometimento conjunto é criação de todas as partes envolvidas, rescindindo apenas com a concordância de todos” (2004, p. 100). Suponha que estudantes decidam fazer um trabalho para a faculdade

juntos. Neste caso, há um comprometimento conjunto; se algum dos estudantes decidir não fazer a sua parte no trabalho, estaria violando o comprometimento, o que constituiria uma ofensa a cada um dos membros do grupo.

A condição de “acreditar em p como um corpo”, por sua vez, não deve ser entendida como uma exigência de que cada um dos membros de um grupo acredite pessoalmente que p , mas que, em contextos apropriados, os membros se comportem de tal forma como se acreditassem que p . Por exemplo, em situações onde a crença que p for questionada, os membros do grupo deveriam defender a posição do grupo quanto a p . No caso apresentado anteriormente, do comitê de contratação, por exemplo, um membro do comitê, ao falar sobre a decisão do grupo sobre a melhor candidata, deveria expressar a posição do grupo, de que Maria é a melhor candidata, independentemente de sua crença pessoal nisso. O componente contextual aqui desempenha um papel relevante, uma vez que não é em todos os contextos que alguém age como membro representante de um grupo. Mas nas situações contextuais relevantes, onde alguém age ou fala pelo grupo, o comprometimento conjunto de acreditar em p como um corpo demanda a obrigação de agir como se acreditasse que p .

A análise de crenças coletivas de Gilbert, no entanto, sofre algumas de dificuldades. A análise de crença coletiva na abordagem de comprometimento conjunto é acusada de circularidade (TUOMELA, 1992), uma vez que a noção de crença coletiva parece ser pressuposta na condição de acreditar em p como um corpo. A maior crítica, no entanto, é que Gilbert não oferece um modelo de crença coletiva, mas de *aceitação* coletiva (WRAY, 2001). O conceito de “aceitação” pode ser caracterizado como uma atitude doxástica relativa a uma proposição p , mas com pelo menos duas diferenças importantes da atitude de crença que p . Primeiro, aceitações estão sob controle da vontade, enquanto crenças são tipicamente tratadas como atitudes involuntárias: eu não posso simplesmente decidir deixar de acreditar que estou ao notebook escrevendo este texto. Aceitações, por outro lado, parecem admitir tal controle: eu posso aceitar – para fins de discussão sobre o ceticismo cartesiano, por exemplo – que não estou ao notebook neste momento, mas estou sendo enganado por um gênio maligno que me faz acreditar estou. Segundo, enquanto crenças objetivam a verdade, aceitações tipicamente são associadas a algum tipo de utilidade ou fim pragmático: *acreditamos* em uma proposição p , porque consideramos que p é verdade; deixaríamos de acreditar em p , se algo indicasse que p é falso. No

entanto podemos aceitar que p , mesmo quando não consideramos que p é verdade— eu não considero que a hipótese do Gênio Maligno seja verdadeira, mas posso aceitá-la, com a finalidade de estudar o ceticismo cartesiano. Como a análise de Gilbert de crenças coletivas parece permitir que grupos escolham sua atitude com relação à proposição em questão, a atitude coletiva analisada nos termos de comprometimento conjunto parece ser aceitação e não crença. Essas considerações fomentam o debate entre crentes (aqueles que defendem que crenças coletivas são crenças genuínas) e rejeicionistas (aqueles que defendem que crenças coletivas são aceitações coletivas). O debate entre crentes e rejeicionistas, no entanto, é um debate sobre ontologia coletiva e não sobre epistemologia coletiva, embora tenha implicações epistemológicas. Por exemplo, seria esse um problema para a tese de que há sujeitos epistêmicos coletivos? Há em princípio duas alternativas para o epistemólogo coletivo responder a essa questão. A primeira seria aceitar que conhecimento também pode ser aceitação verdadeira justificada, tal estratégia é admitida por Tuomela (2004) e Hakli (2007), em análises de conhecimento coletivo. Assim, grupos poderiam ser considerados sujeitos epistêmicos coletivos, a despeito de suas atitudes doxásticas coletivas serem aceitações e não crenças. A segunda seria insistir que crenças coletivas são crenças genuínas, e não aceitações coletivas, conforme argumenta Tollefsen (2003). Seja qual for a alternativa, portanto, há espaço para sustentar a tese de que grupos podem ser sujeitos epistêmicos coletivos e questionar o pressuposto do individualismo epistêmico₂.

Questões para discussão

1. Imagine que, ao chegar do trabalho, João descobre que seu apartamento pegou fogo. Chocado com as perdas, João desabafa com um amigo: “eu sei que meu apartamento pegou fogo, mas não acredito nisso!” Seria esse um contraexemplo à necessidade da crença na análise de conhecimento? É possível alguém saber que p sem acreditar que p ?
2. Como a epistemologia do testemunho lidaria com a existência de *fake news*? A existência de notícias falsas seria uma razão em favor do reducionismo ou um anti-reducionista conseguiria explicar por que alguém não está justificado em acreditar em *fake news*, sem renunciar à tese de que o testemunho é uma fonte legítima de justificação?

3. Se é possível que um grupo acredite que p , enquanto a maioria dos membros de G acredite pessoalmente que não- p , poderia a crença coletiva que p ser justificada? Ou o fato de que maioria dos membros do grupo acredita que não- p constitui uma razão contrária para o grupo acreditar coletivamente p ?

Leitura recomendada

O Livro *Conhecimento e justificação: problemas de epistemologia contemporânea* (LUZ, 2013) trata de maneira introdutória e acessível problemas relacionados à análise do conhecimento e às teorias da justificação epistêmica. Os volumes *Epistemologia Social: Dimensão Social do Conhecimento* (MÜLLER; RODRIGUES, orgs., 2012), *O que NÓS conhecemos? Ensaio em epistemologia individual e social* (MÜLLER, LUZ, orgs., 2015), e *Ensaio sobre epistemologia do testemunho* (MÜLLER; ETCHEVERRY, orgs., 2017) oferecem uma variedade de artigos introdutórios e avançados sobre problemas em epistemologia social.

Referências

- BONJUOR L. (1973). *The Structure of Empirical Knowledge*. Harvard University Press.
- COADY, C.A.J. (1973). "Testimony and Observation." *American Philosophical Quarterly* 10:149-155.
- _____. (1992). *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
- FRICKER, E. (1987). "The Epistemology of Testimony." *Proceedings of the Aristotelian Society Supplement* 61:57-83.
- _____. (1994). "Against Gullibility," in MATILAL; CHAKRABARTI (eds). *Knowing from Words. Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science)*, vol 230. Springer, Dordrecht.
- _____. (1995). "Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony," *Mind* 104:393-411
- _____. (2002). "Trusting Others in the Sciences: a priori or Empirical Warrant?," *Studies in History and Philosophy of Science* 33:373-83.
- _____. (2004). "Testimony: Knowing Through Being Told," in I. Niiniluoto, et al (eds), *Handbok of Epistemology*. New York: Springer.

- GETTIER, E. (1963). "Is Justified True Belief Knowledge?" *Analysis* 23:121-123.
- GILBERT, M. (1987). "Modelling Collective Belief". *Synthese*, v. 73, n. 1, p. 185–204, 1987.
- _____. (1989). *On Social Facts*. Routledge.
- _____. (1994). "Remarks on Collective Belief". In SCHMITT, F. F. (Org.). *Socializing Epistemology: The Social Dimensions of Knowledge*. Rowman and Littlefield, p. 235–56.
- _____. (2002). "Belief and Acceptance as Features of Groups". *Protosociology*, v. 16, p. 35–69.
- _____. (2004). "Collective Epistemology". *Episteme*, v. 1, n. 2, p. 95–107, 2004.
- GOLDBERG, S., e HENDERSON, D, (2005). "Monitoring and Anti-Reductionism in the Epistemology of Testimony," *Philosophy and Phenomenological Research* 72:600-17.
- GOLDMAN, A. (1979). "O que é Crença Justificada?". *Crítica*. Disponível em <https://criticanarede.com/justificacao.html>
- GRAHAM, P. J. (2004). "Metaphysical Libertarianism and the Epistemology of Testimony," *American Philosophical Quarterly* 41:37-50.
- _____. (2006). "Testimonial Justification: Inferential or Non-Inferential?" *The Philosophical Quarterly*, 56: 84-95
- GREEN, C. R. (2006). *The Epistemic Parity of Testimony, Memory, and Perception*. Ph.D. dissertation, University of Notre Dame.
- HAKLI, R. (2007). "On the Possibility of Group Knowledge Without Belief. *Social Epistemology*", v. 21, n. 3, p. 249–266.
- HUME, D. (1972). *Investigação Acerca do Entendimento Humano*. São Paulo: Companhia Editora Nacional
- LACKEY, J. (2006). "It Takes Two to Tango: Beyond Reductionism and Non-Reductionism in the Epistemology of Testimony," in LACKEY e SOSA, eds. *The Epistemology of Testimony*. Oxford: Oxford University Press.
- LEWIS, D. (1969). *Convention: A Philosophical Study*. Harvard University Press.
- LOCKE, J. (1975). *An Essay concerning Human Understanding*. Oxford: Clarendon Press.
- LUZ, M. A. (2013). *Conhecimento e justificação: problemas de epistemologia contemporânea*. Pelotas. Disponível em <http://nepfil.ufpel.edu.br/publicacoes/1-conhecimento-e-justificacao.pdf>
- MATHIESEN, K. (2011). "Can Groups Be Epistemic Agents"? SCHMID, H. B.;

- SIRTES, D.; WEBER, M. (Org.). *Collective Epistemology*. Ontos. p. 20–23.
- MEIJERS, A. (2002). "Collective Agents and Cognitive Attitudes". *Protosociology*, v. 16.
- MÜLLER, F. M.; RODRIGUES, T. V. orgs. (2012). *Epistemologia Social: Dimensão Social do Conhecimento*. Porto Alegre: EdiPUCRS. [disponível em: <https://editora.pucrs.br/Ebooks/Pdf/978-85-397-0176-6.pdf>]
- MÜLLER, F. M.; LUZ, A. M. orgs. (2015). *O que NÓS conhecemos? Ensaios em epistemologia individual e social*. Porto Alegre: EdiPUCRS, Editora Fi. [disponível em: <https://www.editorafi.org/59---felipe-muller>]
- MÜLLER, F. M.; ETCHEVERRY, K. M. orgs. (2017). *Ensaios sobre epistemologia do testemunho*. Porto Alegre: Editora Fi.
- QUINTON, A. (1976). "Social Objects". *Proceedings of the Aristotelian Society*, v. 76, p. 1–27.
- REID, T. (1764). *An Inquiry into the Human Mind on the Principles of Common Sense*. [<https://www.earlymoderntexts.com/assets/pdfs/reid1764.pdf>]
- _____. (2011) *Essays on the Intellectual Powers of Man (Cambridge Library Collection - Philosophy)*. Cambridge: Cambridge University Press
- SANTO AGOSTINHO (2019), *Patrística - Contra os Acadêmicos, A Ordem, A grandeza da Alma, O Mestre*. Paulos Editora. São Paulo.
- SEARLE, J. R. (1997). *The Construction of Social Reality*. New York: Free Press.
- SELLARS, W. (1997) *Empiricism and the Philosophy of Mind*. Harvard University Press. Cambridge, Massachusetts.
- SHIEBER, J. (2009) "Locke on testimony: A reexamination". *History of Philosophy quarterly*, Volume 26, Number 1.
- STEUP, M. (2004). "Internalist Reliabilism." *Philosophical Issues*, 14: 401–425
- TOLLEFSEN, D. (2002) "Challenging Epistemic Individualism". *Protosociology*, v. 16.
- _____. (2003). "Rejecting Rejectionism". *Protosociology*, v. 18.
- TUOMELA, R. (1992). "Group Beliefs". *Synthese*, v. 91, n. 3, p. 285–318.
- _____. (2004). "Group Knowledge Analyzed". *Episteme*, v. 1, n. 2, p. 109–127.
- WELBOURNE, M. (1981). "The Community of Knowledge," *Philosophical Quarterly* 31:302-14.
- WRAY, K. B. (2001). "Collective Belief and Acceptance". *Synthese*, v. 129, n. 3, p. 319–33.

4

Ética

Bruno Aislã Gonçalves dos Santos
Rafael Martins

1. Introdução

Neste capítulo, pretendemos apresentar em linhas gerais e, por vezes de modo breve, as principais questões que permeiam a área filosófica denominada Ética. Grosso modo, “Ética” é apenas um nome para uma disciplina da filosofia, mas também pode referir outra coisa como, por exemplo, um conjunto de regras a serem seguidos. Nosso ponto aqui é apresentar as quatro grandes subáreas da disciplina filosófica Ética: (a) a Axiologia; (b) a Metaética; (c) a Ética Normativa e; (d) a Ética Aplicada. Nosso intento não é fazer uma apresentação exaustiva das teorias e dos problemas de cada subárea, mas apresentá-las de modo introdutório. Nós começamos, portanto, discutindo o que é Ética, afinal? E tentamos distinguir a área da **Ética** das outras áreas e quais são as dificuldades de oferecer uma definição clara do que seja Ética. Depois, passamos a explicar de maneira geral do que trata cada subárea da disciplina de Ética.

Nesse sentido, começaremos a expor o que é a **axiologia** e nos concentraremos em apresentar três modelos teóricos acerca da natureza da felicidade/bem-estar e de seu valor intrínseco. Nesse ponto, iremos abordar três modelos teóricos que intentam explicar o que é o bem-estar: (a) o Hedonismo, (b) a Teoria da Satisfação das Preferências e (c) a Teoria da Lista Objetiva. A **Teoria Hedonista** sustenta que, grosso modo, nosso bem-estar é composto por estados mentais como o prazer. Já a **Teoria da Satisfação das Preferências** defende que nosso bem-estar é composto pela satisfação de nossas preferências informadas. Por fim, a **Teoria da Lista Objetiva** defende que nosso bem-estar é composto por estados objetivos do mundo, ou seja, há coisas valiosas em si que

compõem nosso bem-estar e não depende de produzir em nós estados mentais.

Depois, apresentamos a taxonomia das **teorias metaéticas**, assim como os principais problemas da área. Em metaética estamos preocupados em responder os problemas mais gerais acerca da moralidade. Por exemplo, será que há conhecimento moral? A moralidade expressa fatos? Assim, discutimos brevemente algumas das posições envolvidas nela. Serão apresentadas as teorias com base nos compromissos epistemológicos e metafísicos. Quando falamos de compromissos epistemológicos e metafísicos queremos dizer que uma teoria sustenta (ou não) que a moralidade expressa conhecimento sobre fatos, por exemplo. Teorias que defendem que a ética expressa proposições são **teorias cognitivistas**, as que negam esta tese são **teorias não-cognitivistas**. É importante ressaltar que os compromissos metafísicos também são relevantes. As teorias que defendem que a moralidade expressa algum tipo de fato são **teorias realistas**, já as que negam essa tese são **teorias irrealistas**³³. São exemplos de teorias cognitivistas que vamos explorar: a teoria de Railton (que é realista naturalista), a de Moore (que é realista não-naturalista) e a de Mackie (que é irrealista). Já as teorias não-cognitivistas são defendidas por Ayer, Blackburn e Gibbard (as três teorias são irrealistas).

Em seguida, discutimos os três grandes modelos normativos: a ética das virtudes, o consequencialismo e o deontologismo kantiano. Em **ética normativa** estamos interessados em estabelecer as normas (regras de condutas) que devemos seguir. Em outros termos, estamos preocupados em definir nossos deveres morais. A ética das virtudes vem em primeiro lugar; lá nos focaremos na teoria aristotélica, mas também indicamos alguns dos defensores contemporâneos de tal modelo teórico. Em **ética das virtudes**, defende-se que devemos cultivar nossas virtudes (um traço de caráter) e agir como um agente virtuoso agiria. Posteriormente, passamos a discutir o consequencialismo em duas vertentes, o egoísmo ético e o utilitarismo. A **teoria consequencialista** sustenta que nossas ações são corretas ou incorretas em função daquilo que elas produzem como consequência. Nos focaremos em linhas gerais nas ideias do egoísmo ético e em algumas variações do utilitarismo. O **egoísmo ético** é a tese segundo a qual uma ação é correta se, e somente se, gera as melhores consequências para a pessoa que está tomando as decisões. Em outros termos,

³³O termo “anti-realista” é também usado para falar dessas teorias.

uma ação é correta se ela beneficia egoisticamente aquele que está decidindo. Já o **utilitarismo** é a teoria segundo a qual uma ação é correta se, e somente se, produz as melhores consequências para todos os indivíduos. Assim, as ações são tomadas como corretas se todos os indivíduos são beneficiados por suas consequências. Sobre o utilitarismo é importante ressaltar que há, pelo menos, quatro variações que iremos explorar: (i) O Utilitarismo de Atos; (ii) O Utilitarismo de Atos de Nível; (iii) O Utilitarismo de Regras do Código Moral Vigente e; (iv) O Utilitarismo de Regras do Código Moral Ideal. Cada uma dessas teorias normativas é diferente dado que cada uma aceita uma formulação diferente da teoria consequencialista mais geral. Por fim, em ética normativa, discutimos o modelo paradigmático do deontologismo, qual seja, o deontologismo de Immanuel Kant. O **deontologismo kantiano** sustenta que nossos deveres devem ser definidos através da nossa racionalidade e que nós devemos agir reconhecendo racionalmente tais deveres. Adicionalmente, segundo Kant, a ética nada tem a ver com as consequências das ações, mas das razões que oferecemos para estabelecer um dever.

E, em nosso último tópico, faremos uma apresentação de algumas das principais áreas da Ética Aplicada e os problemas envolvidos nela. A **Ética Aplicada** é focada em entender e oferecer soluções aos problemas relacionados da nossa interação com outros seres humanos, com animais não humanos, com a natureza, com a tecnologia etc. mediados por técnicas e tecnologias criadas por nós. Quando falamos em “aplicação” da ética estamos interessados em entender a complexidade da realidade e como podemos oferecer soluções para os problemas morais que advêm dela. Assim, em Ética Aplicada, estamos interessados em estabelecer, por exemplo, se os animais têm direitos, se temos algum dever em especial para com a Natureza, se podemos vender qualquer coisa, se o aborto é moralmente permissível, etc.

Isso posto, desejamos que este capítulo seja de alguma ajuda aos estudantes iniciantes nesta maravilhosa área que é a Ética.

1.1. O que é ética?

De uma forma ecumênica, ética, moralidade e filosofia moral são sinônimos e consistem na investigação sobre *como se deve agir* e na elaboração de *códigos de conduta*. Com o objetivo de decidir sobre o que fazer, em

quaisquer circunstâncias, naturalmente avaliamos objetos concretos e figurados, como estados mentais, ações singulares, condutas e caracteres de indivíduos, grupos e instituições, bem como estados-de-coisas; e a ética procura justamente por proposições que sirvam como critérios de avaliação.

Toda avaliação requer valores, ou seja, critérios, referências ou padrões, com os quais julgamos ou comparamos esses objetos. A *primeira tarefa investigativa* da filosofia moral, portanto, é *axiológica* e aborda quais são esses valores; num segundo momento, perguntamo-nos que tipos de objeto possuem valor, como, por exemplo, estados psicológicos subjetivos ou estados objetivos do mundo. Quando um valor é o objetivo final de uma ação, ou seja, um fim em si mesmo e jamais um meio para qualquer outro fim, o chamamos de valor intrínseco. Igualmente importante é a investigação do valor das ações que, enquanto meios, levam a obtenção de itens de valor intrínseco. Dessa maneira, a *segunda tarefa investigativa* da ética refere-se à elaboração de normas, ou seja, proposições como regras, princípios ou fórmulas que compõem códigos de conduta. Ética *normativa*, portanto, refere-se às teorias sobre quais ações corretamente levam a obtenção ou produção de valor intrínseco. Tudo aquilo que possui valor apenas enquanto meio para um valor intrínseco é considerado como possuindo valor extrínseco ou instrumental.

Dessa forma, a maioria das teorias morais apoia-se numa explicação *teleológica* da agência humana, bem como das instituições. Assim, ações e instituições são entendidas como *meios* que possuem valor extrínseco enquanto visam cumprir determinados *fins, objetivos, planos ou propósitos* que, em si, possuem valor intrínseco. Normalmente, quando avaliamos fins, dizemos que eles são *bons* ou *ruins*, ao passo que avaliamos os meios enquanto *corretos* ou *errados* na medida em que promovem fins bons ou ruins. Por isso a ética as vezes é chamada de teoria da bondade ou teoria da correção, porém, como fica evidente, estas são diferentes partes da investigação moral como um todo, onde a primeira é denominada axiologia ou teoria do valor intrínseco, e a segunda, ética normativa.

Um esclarecimento é necessário. Dizer que teorias morais se apoiam em teorias teleológicas não as compromete com nenhuma teoria moral substancial, como as teorias deontológicas ou consequencialistas, pois a explicação teleológica apenas descreve a estrutura mental e material da ação humana, ou seja, atos concretos produzidos por estados mentais com o objetivo

de atingir uma determinada meta, cumprir um determinado propósito, ou promover um determinado estado-de-coisas. Nesse sentido, a teoria teleológica da ação é moralmente neutra, ou seja, em si, a teoria não atribui maior ou menor significância moral seja aos estados mentais agenciais, seja aos estados-de-coisas almejados ou produzidos, seja ao ato concreto em si aparte suas fontes causais e consequências. Note que as teorias deontológicas também utilizam estruturas teleológicas, mas o foco da avaliação deontológica é na intenção da ação e não em seus resultados. Porém, as teorias deontológicas precisam explicar de um jeito ou de outro o propósito das ações humanas, e ainda que seu valor não derive de seus resultados, é preciso dizer por que o agente agiu da forma como o fez.

Para visualizar como uma avaliação normativa funciona, pense nos programas de televisão *The Voice* e *Masterchef*. No primeiro, temos performances musicais, e no segundo performances culinárias, que são executadas por agentes dotados de capacidade de escolha, erudição, destreza etc. que decidem sobre a melhor forma de agir de acordo com certas normas. Os jurados desempenham, por sua vez, a função de agentes avaliadores, ou seja, examinam em que medida as performances se adequam àquelas mesmas normas. Na esfera normativa musical os jurados empregam critérios como tom, ritmo, composição etc. e, na esfera normativa gastronômica, critérios como corte, sabor, apresentação etc. Essas normas, também são comumente chamadas de padrão, critério, e, no caso da moralidade, valor. Dessa forma, as performances e seus resultados são valiosos na medida em que instanciam esses valores – em outras palavras, as melhores performances são aquelas que melhor instanciam cada um dos valores internos à sua esfera normativa. No caso da ética, podemos avaliar condutas de acordo com valores definidos por nossas teorias axiológicas, para citar apenas alguns, liberdade, igualdade e felicidade. Assim, as ações podem ser valiosas porque são instâncias genuínas de ações livres, igualitárias, hedonistas, ou porque promovem um desses valores.

Uma última distinção preliminar separa a moralidade em descritiva e normativa. A moralidade descritiva refere-se a códigos de conduta adotados e postos em prática por grupos ou sociedades reais (como uma lista de regras afixada na parede de um templo). Enquanto a moralidade normativa refere-se a códigos de conduta que seriam adotados e promovidos por todos os seres *racionais em condições específicas*. No primeiro sentido reconhecemos um sem-

número de códigos adotados por diferentes grupos e sociedades em diferentes lugares e épocas. Kant é um precursor da investigação moral descritiva na última e menos popular de suas obras *Antropologia sob um Ponto de Vista Pragmático*. Porém, Kant é mais conhecido por ter captado o sentido normativo da moralidade melhor do que ninguém. No sentido normativo, a moralidade se refere a um código de conduta que seria aceito por qualquer pessoa que atenda a certas condições epistêmicas e volitivas, quase sempre incluindo a condição de ser racional. Quando uma pessoa atende a essas condições, essa pessoa é reconhecida como um agente moral.

Nem todos os códigos de conduta adotados e postos em prática pertencem à moralidade descritiva, pois existem códigos de conduta não-morais, como códigos de conduta desportiva, profissional, legal, prudencial e assim por diante. Ao mesmo tempo, nem todo código de conduta adotado por agentes racionais é necessariamente moral. Dado que é plausível que seres racionais adotem códigos de conduta não-moral, como por exemplo um código exclusivamente prudencial. Portanto, em ambos os casos, é necessário que encontremos as características definicionais da moralidade, que seriam incluídas em qualquer definição da ética, para que nossa investigação moral seja capaz de diferenciar a esfera normativa moral de outras esferas normativas. Todavia, não há consenso sobre quais são estas características, logo ambas as moralidades descritiva e normativa sofrem do problema de demarcação.

1.2. O problema da definição da ética

Devido a sua abrangência de objetos, diversidade teórica e similaridades com outras esferas normativas, não é fácil definir o campo de investigação da ética. Isso gera o problema de definição da ética. Na base do problema está a distinção entre concepções de moralidade e teorias morais substanciais, paralela à distinção que John Rawls³⁴ traçou entre a concepção

³⁴O conceito de justiça (...) [é] definido, então, pela sua função de [estabelecer] seus princípios na atribuição de direitos e deveres e na definição da divisão apropriada das vantagens sociais. [ou seja, o conceito formal de justiça é definido por sua função de estabelecer direitos, deveres e dividir vantagens sociais.] Uma concepção de justiça é uma interpretação deste papel” [ou seja, qualquer teoria que partilhe de forma

geral de justiça e várias teorias substanciais da justiça. A concepção geral de justiça estabelece *a priori* as propriedades que toda teoria substancial da justiça precisa ter, para ser considerada como tal. Rawls acreditava que, por exemplo, todas as teorias substanciais da justiça têm como forma a função distribuir direitos, deveres e vantagens sociais, diferindo apenas na forma como interpreta esses conceitos e como os distribui. Uma teoria normativa incompatível com essa função não respeita a concepção geral de justiça e, portanto, não é uma teoria da justiça substancial genuína.³⁵

Como vimos acima, existem diversos domínios normativos para se avaliar ações e as teorias morais não são as únicas. Claro que ações podem ser morais ou imorais, mas também podem ser prudentes ou imprudentes, desportivas ou antidesportivas, legais ou ilegais, a assim por diante. Logo, surge a questão: o que distingue a avaliação moral em particular das outras categorias de avaliação? Certamente a prudência, o desporto, a lei, assim como a culinária e a música são categorias de avaliação e, portanto, esferas normativas. Mas o que especificamente distingue a moralidade destas outras esferas normativas? Para responder esta pergunta, seria necessário apontar não apenas qual ou quais as suas características necessárias, mas também características que só a investigação moral possui, pois, do contrário, estaremos identificando características comuns a mais de uma esfera, e não características distintivas. Além disso, estas propriedades agiriam como critérios de entrada, ou seja, uma teoria moral só contará como tal, se possuir tais características distintivas, como no caso da justiça, citado acima.

Mas porque tentar definir os contornos formais distintivos da moralidade preliminarmente? Já não basta dizer que a moralidade se refere a avaliação de ações e códigos de conduta? Uma razão popular é a aparente necessidade de se estabelecer critérios de participação. Para se entender essa necessidade, note que para determinarmos qual das teorias morais rivais é a correta,

substancial direitos, deveres, e vantagens sociais qualifica-se como uma teoria de *justiça*. Portanto, para uma teoria substancial ser qualificada como teoria da justiça, ela precisa *distribuir direitos, deveres e vantagens*.] John Rawls, "Theory of Justice", p. 9.

³⁵Inclusive Rawls utiliza esta distinção como parte de seu argumento de que o utilitarismo não é uma teoria substancial da justiça genuína, por não exibir algumas características formais que, segundo ele, deveriam configurar qualquer concepção de justiça.

precisamos testar a coerência dessas teorias contra as nossas convicções morais mais seguras sobre o que fazer em tipos de casos mais ou menos específicos. Mas para desempenhar esse teste, precisamos distinguir nossas convicções morais de outros tipos de convicção. Além disso, ao tentar determinar a teoria moral correta, nós também comparamos várias teorias normativas substanciais entre si e seus veredictos particulares sobre o caso em análise. Porém nós não testamos teorias morais contra juízos que não possuem conteúdo moral. Por exemplo, nós não comparamos uma conclusão moral substancial qualquer sobre aborto com nossas convicções culinárias ou musicais. Para não incorreremos nesse erro, precisamos saber o que distingue um juízo moral de um juízo sem conteúdo moral, pois do contrário chegaríamos a resultados normativos distorcidos e não genuinamente morais. Por isso, precisamos de uma definição de ética preliminar que atue como critério de entrada, ou o projeto de investigação moral não poderia sequer começar.

Quais características formais inerentes a toda teoria moral substancial poderiam desempenhar esse papel? Existem muitas teorias sobre a distinção da moralidade e, logo, sobre qual seria a correta definição de ética. Para o propósito deste capítulo podemos nomear: a teoria do conteúdo moral (CM), a teoria da justificação moral (JM), a teoria reativa-attitudinal (RA), a teoria da motivação moral (MM) e a teoria do prevailecimento racional da ética (PR).

De acordo com CM, é o conteúdo substancial da ética que lhe provê sua distinguibilidade. Teóricos-CM acreditam que o conteúdo da moralidade é sempre preenchido por avaliações sobre ações estritamente públicas ou interpessoais em oposição a ações estritamente privadas ou intrapessoais, que por sua vez podem pertencer à esfera da prudência, da estética, da culinária e assim por diante. Ou seja, o conteúdo da moralidade é sempre relacionado aos interesses de todas as pessoas, menos o agente. No sentido normativo, o conteúdo da moralidade são obrigações que devemos uns aos outros, mas não a nós mesmos (pois estas seriam obrigações pertencentes a outros domínios normativos). Um primeiro problema aqui é que seria difícil decidir se uma infinidade de ações é genuinamente interpessoal ou meramente intrapessoal. Note que uma ação estritamente intrapessoal poderia até produzir externalidades ³⁶ morais. Mas seria difícil defender que um agente seja

³⁶ Consequências não intencionais ou não desejadas.

penalizado por estas externalidades morais se, por exemplo, elas estiverem fora do alcance epistêmico do agente, ou estiverem fora de seu alcance. Então, a distinção entre ações interpessoais e intrapessoais força a adução de um critério de avaliação de externalidades, como, por exemplo, a teoria da responsabilidade negativa³⁷. Mas obviamente, muitas teorias morais, como por exemplo a doutrina Kantiana, não possuem esse critério e ainda assim são consideradas teorias morais plenas. Logo a distinção não cumpre uma das suas principais funções que é atuar como critério de entrada.

Mas CM enfrenta outros problemas. Mesmo se a distinção entre ações intrapessoais e interpessoais for clara, o que diferiria a ética de outras esferas avaliativas que também se ocupam de ações interpessoais ou do interesse alheio? A medicina por exemplo se ocupa muito mais largamente dos interesses dos pacientes do que dos interesses dos médicos. Muitos outros domínios avaliativos se ocupam com os interesses dos outros que não o agente. Por exemplo, o alvo da estética é as vezes muito mais centralizado no interesse do apreciador. Mas então o que diferencia interesses morais de interesses médicos ou interesses estéticos? À primeira medida, seria então diminuir a variedade de esferas avaliativas concentradas em ações interpessoais. A lista é longa, poderíamos citar normas feudais, normas desportivas, normas de etiqueta, e assim por diante. Portanto, a única coisa que a teoria CM faz é identificar uma serie de domínios avaliativos interpessoais, porém independentes, onde a moralidade é apenas um membro, cuja diferença para com os outros ainda não sabemos. Aqui o projeto de definição da ética não cumpre seu principal papel que é o de estabelecer um critério de entrada para fins de investigação moral substancial.

Isso ocorre porque a barreira de entrada é baixa demais pois permite a participação de teorias normativas não-morais na investigação moral. Mas ela também pode ser alta demais, ao excluir teorias morais que, como a de Kant, argumentam que a moralidade também governa a conduta que afeta exclusivamente o agente. Baseado na sua concepção particular de racionalidade, Kant defende que temos obrigações morais intrapessoais, como por exemplo, desenvolver nossas capacidades intelectuais individuais completamente

³⁷Bernard Williams, *Utilitarianism: For and Against*, 2008, Cambridge University Press, p. 93.

independente de suas consequências interpessoais. Se a barreira de entrada for alta demais, teorias morais bastante plausíveis, pelo menos inicialmente, ficam de fora da investigação moral substancial sobre casos concretos. Um outro exemplo de como a interpessoalidade pode ser uma barreira de entrada muito alta é a questão sobre o que devemos fazer com os mortos. Muitas teorias morais interpessoais são baseadas no valor da racionalidade ou do bem-estar das pessoas. Ou seja, essas teorias geram obrigações morais de acordo com esses valores. Porém, os mortos não são racionais, nem possuem bem-estar, não são pessoas. E, ainda assim, muitos de nós acreditamos ter obrigações morais em relação aos mortos, ou seja, obrigações não-pessoais e ainda assim morais. Portanto definir a interpessoalidade como conteúdo da ética excluiria qualquer teoria moral que, baseada num valor qualquer, gere obrigações em relação aos mortos.

Note que não estamos dizendo que não há possível definição da ética ou que não existem verdades formais ou *a priori* sobre a moralidade. Provavelmente existem, mas não podemos dizer que CM é uma delas, dado que ela sofre do problema de ser um critério de entrada muito ruim.

Vejamos outras possibilidades. Para JM o que distingue os juízos morais de outros tipos de juízos normativos é o tipo de coisa que pode ser mobilizado como justificação destes juízos. Embora exista enorme dissenso sobre que tipo de coisas podem atuar como justificação de juízos morais, pode ser que a justificação moral seja distinta de todas as outras esferas normativas, como as citadas acima por exemplo, justamente nesse aspecto.

Que tipos de coisas podem justificar proposições morais? Temos por exemplo métodos justificatórios baseados em contrato hipotético, natureza da racionalidade humana ou autonomia, perspectiva da segunda-pessoa, psicologia emocional, platonismo conceitual, para dizer os mais populares. Porém perceba que qualquer outro domínio avaliativo pode ser justificado da mesma maneira. A psicologia emocional explica nossos juízos estéticos e o platonismo conceitual é utilizado em matemática. Pior ainda para JM é que podemos fazer ética normativa a partir de uma posição completamente agnóstica da perspectiva justificatória. Mas perceba que fazemos ética normativa, ou seja investigação moral substancial, sem ter certeza última sobre qual é a mais fundamental justificação das proposições morais. Podemos estar incertos sobre se o que realmente justifica a incorreção do aborto são considerações sobre a estrutura

senciente dos envolvidos ou considerações sobre sua autonomia, e nem por isso deixamos de fazer investigação moral substancial e tomar decisões sobre aborto. Mais fácil de entender é que não precisamos estar certos sobre que tipo de coisa justifica um juízo moral, para emitirmos um juízo moral, assim como inicialmente não sabemos o que justifica o céu ser azul quando concluímos que o céu é azul. Portanto JM não é uma boa teoria da distinção da ética. Portanto, a JM não é suficiente para distinguir a moralidade.

Para a RA, proposições morais são justificadas em função das reações emocionais negativas ou positivas que temos, ao vivenciar situações de todos os tipos – negativa quando, por exemplo, vemos alguém jogar lixo no chão, ou positiva frente a atos de filantropia desinteressada. Novamente, veja que essas reações emocionais ocorrem em qualquer outra ocasião, em qualquer outro contexto. Claro que nossas reações emocionais são as piores possíveis quando vemos nosso time favorito perder uma final de campeonato, mas isso não justifica qualquer proposição moral. O mesmo com a estética. Portanto. RA tem um problema de subdeterminação extrema, ou seja, seu critério de entrada é muito leniente e confunde milhares de domínios avaliativos, deixando a investigação moral substancial praticamente inviável.

MM, por sua vez, credita a distinguibilidade da ética à sua força motivacional especial. Segundo MM, nenhum outro domínio normativo possui uma conexão tão forte entre um juízo e a motivação correspondente como a moralidade. A teoria MM sobre a distinção da moralidade argumenta que a moralidade, diferentemente de outras esferas normativas, é a única cujos juízos necessariamente motivam o agente que os endossa a agir de acordo. Se nós acreditamos que doar nossas roupas sem uso para os desabrigados é bom, então, necessariamente nós estamos motivados a fazê-lo, ou então alguma coisa está errada conosco. Podemos estar mentindo para nós mesmos para nos sentirmos melhores, ou podemos ter alguma desordem mental que nos ceifa a vontade de viver. Mas em condições normais de racionalidade, pessoas normalmente estão mais fortemente motivadas a agir de acordo com suas convicções morais do que de acordo com outras convicções normativas. Por exemplo, imagine que eu seja um ávido apreciador de jazz. Fico sabendo que hoje à noite haverá uma apresentação de jazz perto da minha casa. Eu estou convicto que eu devo ir a esse show. Claro que estou motivado para ir ao show. Mas também estou convicto que devo terminar a minha série favorita. A noite

está fria. Resolvo ficar em casa. Mas agora imagine que no caminho para sua entrevista de emprego, você se depara com uma criança se afogando num lago próximo a você. Claro que você está motivado para sua entrevista, muito motivado na verdade. Porém, é muito mais forte a sua motivação correspondente à sua convicção de que é seu dever salvar a criança. Você perde a entrevista e pula na água. Porém outros juízos normativos igualmente motivam agentes. Posso me sentir igualmente motivado a agir de acordo com meus juízos prudenciais, bem como de acordo com meus juízos estéticos, e assim por diante. Portanto a última tentativa de refinar MM é dizer que a moralidade possui uma força racional maior do que todos os outros domínios normativos.

Repare que em todos os momentos em que uma decisão é requerida, avaliamos a situação de acordo várias perspectivas normativas, como prudência, moral, estética, profissional, político, paternal, e assim por diante. Cada uma dessas esferas avaliativas gera razões para agir, num sistema de *input/output*, onde o input é o objeto avaliado e o output é a razão para agir. Dessa forma, em todo momento em que uma decisão é requerida temos diversas razões, por exemplo estéticas, morais, prudenciais, legais, políticas, civis, familiares e assim por diante, “competindo” pela ação do agente, ou seja, aquela que vencerá a decisão tomada. Mas de acordo com PR as razões morais *sempre* vencem todas as outras razões, até mesmo as razões prudenciais mais rigorosas como a própria sobrevivência. Aristóteles ilustra PR com eloquência:

Ainda que tal fim [bem-estar] seja o mesmo tanto para o indivíduo como para a cidade, o deste último parece ser algo maior e mais completo, quer a atingir, quer a preservar. Embora seja valioso atingir esse fim para um indivíduo só, é mais belo e mais divino alcançá-lo para a cidades-Estados.(ARISTÓTELES, 1094b)

Para ver como PR é rigorosa, imagine uma situação onde a salvação da cidade depende exclusivamente de você. Suponha que você, ao se lançar em salvamento da cidade, mesmo com sucesso, terá sua vida sacrificada. De acordo com PR, não resta dúvida sobre o que deve ser feito. Agora suponha que existe a possibilidade de você viver em segurança em outra cidade. Será que não seria melhor ser prudente e viver em segurança? Bem esta é uma questão

que só uma investigação normativa pode decidir. Mas agora está claro como PR é rigorosa. PR atribui a especialidade da ética à sua força normativa superior, ou seja, via de regra as exigências morais vencem exigências geradas por outras esferas normativas, independentemente de as exigências serem justificadas solidamente dentro de suas respectivas esferas.

Entretanto, não há como saber se as razões morais realmente *sempre* prevalecerão em todos e quaisquer casos. É plausível que existam pelo menos alguns domínios não morais, cujas exigências nos dão razão não-moral para agir. O caso acima ilustra a possibilidade de razões prudenciais prevalecer sobre razões morais em alguns casos. Assim como não é difícil imaginar casos em que a lei não nos permite agir de acordo com sólidas razões morais. Portanto, PR não é uma boa teoria da distinguibilidade da ética, pois não é plausível dizer que razões morais *sempre* prevalecem em casos de conflito com razões originais de domínios normativos concorrentes.

A lição que fica é que definições de moralidade não podem, de partida, ser usadas como argumentos finais contra teorias morais inicialmente plausíveis. Dito isso, é muito difícil dizer que uma única definição de ética encamparia todas as teorias morais consagradas, pois é plausível considerar que nem todo agente racional *necessariamente* optaria sempre por agir moralmente em casos de conflito normativo. Além disso, teorias morais substanciais especificam diferentes conjuntos de condições necessárias para decisões racionais serem tomadas. Sabemos que moralidade se refere à investigação e à avaliação de diferentes objetos e códigos de conduta.

Devemos abandonar o projeto de delinear os contornos distintos da esfera moral? Nossa sugestão é inverter a direção conceitual. Até o momento vimos que a direção conceitual parte da forma da moralidade para a moralidade substancial, ou seja, estabelece de antemão quais teorias avaliativas substanciais são teorias morais a partir de critérios formais estabelecidos por uma concepção geral de ética, por sua vez definida por características distintivas. A direção conceitual inversa implica numa investigação do conteúdo substancial da esfera moral, ou seja, uma investigação do conteúdo das razões, demandas, e considerações morais. Quando soubermos quais são estas razões, demandas e considerações, seremos capazes de determinar as características formais inerentes ao domínio da investigação moral.

1.3. As Subdivisões da Filosofia Moral

No âmbito acadêmico, convencionou-se dividir a filosofia moral em 1. Axiologia; 2. Metaética; 3. Ética normativa e 4. Ética aplicada.

A **axiologia**, também chamada de **teoria do valor intrínseco**, investiga exatamente no que consistem os valores fundamentais e qual o valor das coisas, seja ele positivo, negativo ou neutro. Neste âmbito falamos em valores extrínsecos ou instrumentais e valores intrínsecos. Tudo que possui valor extrínseco possui valor apenas e na extensão em que é um meio ou instrumento para se obter valor intrínseco. Habitualmente, discute-se, em teorias do valor intrínseco, as concepções de bem-estar, mas podemos também discutir se seres humanos possuem valor intrínseco ou, ainda, se a Natureza possui valor intrínseco (esta última discussão é muito popular em ética ambiental) – no ponto 2 discutimos com algum pormenor as teorias sobre a natureza do bem-estar. Assim sendo, a Axiologia está intimamente ligada à Metaética. Mas do que se trata esta última?

A **metaética** é uma área central da Ética, assim como a ética normativa e a ética prática ou aplicada. Como qualquer área da Filosofia que inclui o prefixo “meta” em seu nome, os estudiosos da metaética lidam com os problemas mais gerais pertencentes à Ética: são problemas metafísicos, epistemológicos, semânticos etc., que fazem parte da grande área da Ética.

Enquanto que a ética normativa discute, digamos, o problema de saber se uma ação boa é aquela que produz boas consequências, a metaética lida com o problema (muitíssimo mais geral) de saber se há uma análise bem-sucedida do conceito de bom. Enquanto que a ética prática estuda, digamos, o problema de saber se a afirmação “o aborto é moralmente permissível” é verdadeira, a metaética estuda o problema de saber se as afirmações morais têm valor de verdade. Assim, como observamos, a metaética lida com os problemas mais gerais da Ética (SANTOS & MERLUSSI, 2013, p. 347).

Então, em metaética, tratamos, por exemplo, de problemas da semântica moral, da metafísica moral, da epistemologia moral, da fenomenologia moral, da psicologia moral etc. E do que trata a ética normativa, então?

A **ética normativa** trata do estudo dos critérios, procedimentos, princípios, modos de ser que nos permitam separar e avaliar aquilo que é bom e/ou correto daquilo que é mau/incorreto. Há três grandes modelos teóricos em ética normativa, quais sejam: (a) **Éticas das virtudes**: defendem, grosso modo, que o que devemos fazer em uma dada circunstância é o que uma pessoa virtuosa faria; (b) **Ética deontológicas**: defendem que há um princípio ou norma cuja validade é autoevidente ou justificada pela razão e que devemos seguir tal norma, já que ela revelaria o que é moralmente correto/bom e incorreto/mau, por fim; (c) **Éticas consequencialistas**: são aquelas teorias que defendem que devemos produzir as melhores consequências – em outros termos, uma ação, regras, instituição, motivo, etc. é correto/bom ou incorreto/mau dependendo das consequências que gera. Temos duas grandes teorias que assumem o consequencialismo em seu corpo teórico, quais sejam, (c1) **O egoísmo ético**³⁸: tese segundo a qual uma ação é correta/boa se, e somente se, gera as melhores consequências para o tomador de decisão e; (c2) **o utilitarismo**: tese segundo a qual uma ação é correta/boa se, e somente se, gera as melhores consequências para todos os indivíduos de forma imparcial. As melhores consequências são aquelas que geram o máximo de bem-estar.

Por fim, temos a **ética aplicada** (por vezes chamada de ética prática) que, como o nome sugere, lida com a aplicação das normas e princípios da ética normativa a problemas existentes na realidade. É um tanto quanto comum dividir a ética aplicada em algumas áreas como, por exemplo: (i) a **Bioética**; (ii) A **Ética Ambiental**; (iii) A **Ética Animal**; (iv) A **Ética Econômica**; (v) A **Infoética** (vi) As **Éticas Profissionais** etc. Esta subdivisão não é estanque nem exaustiva, pode haver mudanças dado as nossas transformações sociais e o aparecimento de

³⁸ Importante notar que, quando falamos de egoísmo, temos que nos atentar para diferença entre egoísmo ético e egoísmo psicológico. O egoísmo ético é uma teoria normativa, ou seja, ela nos diz o que fazer. No caso do egoísmo ético devemos fazer o que é melhor para nós. Já o egoísmo psicológico é uma teoria descritiva sobre a motivação humana. A teoria sustenta que as pessoas são sempre motivadas por razões egoístas. Aqui estamos nos ocupando da primeira e não da segunda.

novas tecnologias e as mudanças das relações entre seres humanos e o mundo.

Apesar da grande área da Ética se dividir nestas quatro subáreas há uma ligação profunda entre elas. Mas, não se engane, manter separados os problemas que cada subárea trata não é apenas um expediente de organização da disciplina de Ética. É um expediente de organização dos nossos pensamentos e argumentos. Afinal, você pode oferecer bons argumentos metaéticos para sustentar algo dentro da ética normativa, assim como oferecer argumentos acerca da aplicabilidade de um modelo normativo para tentar mostrar a plausibilidade de tal modelo. Todavia, não misture as coisas, os problemas de cada subárea demandarão instrumentais teóricos distintos, apesar de serem inter-relacionados.

Considerada esta distinção em subáreas, iremos abarcar em linhas gerais o que há de mais relevante em cada uma delas. Começamos pela axiologia e iremos discutir os três principais modelos de como entendemos o bem-estar. Depois, consideramos, em linhas gerais, as principais teorias metaéticas. Seguindo, iremos considerar os três grandes modelos normativos: a ética das virtudes, o consequencialismo e o deontologismo. Por fim, faremos algumas considerações sobre ética aplicada, indicando algumas de suas subáreas e os problemas pertencentes a elas.

2. Axiologia e Teorias do Valor Intrínseco³⁹

Desde a obra *Ética a Nicômaco*, de Aristóteles, filósofos e filósofos passaram a aceitar que há três classes de coisas que podem nos beneficiar. Há (i) aquelas coisas que podem nos beneficiar apenas enquanto instrumento. Essas coisas *servem* para que consigamos outras coisas. Por exemplo, o dinheiro é apenas um instrumento para que consigamos alcançar outras coisas. Dessa forma, coisas do tipo (i) apenas têm valor se nós lhe atribuirmos algum valor, por isso diz-se que coisas como dinheiro têm valor extrínseco. Há também uma segunda classe de bens: (ii) aquelas coisas que podem nos beneficiar enquanto são utilizadas como instrumentos, mas que também são valiosas por si

³⁹Esta parte do capítulo é uma versão do artigo de SANTOS, B.A.G. Três Teorias sobre o Bem-estar. *In*. Crítica. London, 2017. Disponível em https://criticanarede.com/fil_felicidade.html

mesmas. Por exemplo, o conhecimento é algo que é bom por si mesmo, uma vez que nos faz viver na verdade e não na falsidade, mas também nos serve para fazer outras coisas, como, por exemplo, curar doenças, construir casas, entender o mundo etc. E, por fim, há (iii) coisas valiosas por si mesmas, ou seja, que são boas e que não servem como instrumentos para nós, ou seja, que possuem somente valor intrínseco. Quando nos perguntamos o que é o bem-estar, geralmente estamos procurando quais as coisas que o constituem. Mas também estamos interessados em saber se o próprio bem-estar é um bem do tipo (iii). Aqui, nos preocuparemos em expor as teorias que estão interessadas em responder o que constitui o bem-estar. Assim, sempre que nos perguntarmos o que é o bem-estar, estamos preocupados em saber quais tipos de bens (e que bens) que o constituem e nos beneficiam não instrumentalmente, ou seja, aquilo que possui valor intrínseco. É relevante sabermos o que é o bem-estar, porque há muitas teorias morais que dependem de tal concepção.

Para algumas teorias normativas a relação entre o que é o bem-estar e o que devemos fazer é bem estreita. Por exemplo, para um utilitarista⁴⁰, a força “justificacional” de todo e qualquer raciocínio moral recai sobre o bem-estar. Mas, o que é bem-estar? Há, pelo menos, três teorias amplas acerca do bem-estar. Parfit (1984) dividiu as três principais teorias acerca do bem-estar e seu valor do seguinte modo: (a) teorias da experiência (sendo a principal o hedonismo); (b) teoria das preferências (sendo a principal a teoria das preferências informadas) e; (c) teorias objetivas (sendo a principal a teoria da lista objetiva). Cada uma delas defende um tipo de propriedade que faz do bem-estar valioso por si mesmo e que identifica do que o bem-estar é composto. Teorias como o hedonismo e a das preferências informadas são teorias subjetivistas a respeito da concepção de bem-estar. Já teorias como a teoria da lista objetiva podem aceitar vários elementos como componentes do nosso bem-estar, mas defendem que aquilo que compõe nosso bem-estar não precisa nos afetar subjetivamente⁴¹.

⁴⁰Relembre que utilitaristas defendem que uma ação é moralmente correta se, e somente se, gera como consequência a maior soma de bem-estar para todos os envolvidos de forma imparcial.

⁴¹Apesar de a teoria da lista objetiva poder ser objetivista quanto a concepção de bem-estar, ela também pode ser mista, mesclando tanto elementos subjetivistas quanto elementos objetivistas que compõem o bem-estar, como defendi em Santos (2015).

Teorias hedonistas são subjetivistas no tocante a que elas defendem que o que faz da nossa vida valiosa é a fruição de estados mentais prazerosos – mais sobre isso adiante. Hedonistas explicam, geralmente, o valor dos estados mentais recorrendo a uma propriedade que faz deles bons, qual seja a propriedade de ser “aprazível”. Então, é uma condição necessária e suficiente para que algo seja bom para nós a instanciação da propriedade de ser aprazível. Apenas estados mentais que instanciam tal propriedade podem possuir valor intrínseco. Todo o resto (atividades, estado de coisas, capacidades, traços de caráter etc.) possuem valor extrínseco (ou seja, não tem valor por si mesmo) enquanto provocam em nós tais estados mentais. Já teorias da lista objetiva, por exemplo, são objetivistas, pois defendem que o que faz da nossa vida valiosa são estados de coisas objetivos, independentemente de se eles causam em nós estados mentais prazerosos. Por exemplo, se a beleza é boa por si mesma, então não importa se há alguém para desfrutar de estados prazerosos de sua fruição. Assim algo é bom por si mesmo, ou seja, possui valor intrínseco, se instancia a propriedade de ser belo. Sendo assim, as teorias hedonistas podem ser consideradas subjetivas e puramente pessoais, enquanto as teorias da lista objetiva podem ser consideradas puramente impessoais e objetivas.

Considerando o que foi dito, nesta seção, ofereceremos, primeiramente, uma exposição geral do hedonismo clássico de Bentham e Mill como exemplo de teoria do tipo (i) e indicamos seus problemas. Escolhemos oferecer as teorias de Bentham e Mill, pois elas são casos paradigmáticos de teorias hedonistas. Em um segundo momento, eu ofereçeremos uma explicação da chamada teoria da satisfação de preferência (ou desejos), como exemplo de teoria do tipo (ii) e indicaremos alguns de seus problemas. Faremos uma apresentação geral de tal teoria sem a pretensão de apresentar todas as suas variações. Por fim, apresentaremos a caracterização geral da teoria da lista objetiva, como exemplo de teoria do tipo (iii) e indico seus problemas. Mas antes de passarmos a explorar essas teorias, vamos deixar mais clara a classificação entre teorias subjetivas e objetivas do bem-estar.

2.1. Hedonismo Clássico: Bentham & Mill

Nesta subseção, pretendemos expor a teorias de Bentham e Mill, com a finalidade de apresentar suas principais características e em quais níveis elas se

diferenciam. Isso é importante para que possamos ver claramente o que torna uma teoria sobre o bem-estar, uma teoria hedonista. Também é relevante para entendermos como Mill e Bentham discordam sobre pontos relevantes, dentro de um mesmo tipo de teoria. Sendo assim, num primeiro momento, vamos expor a teoria de Bentham e logo em seguida a de Mill, tentando deixar claras as diferenças entre as duas visões.

Uma das mais antigas teorias acerca do bem-estar⁴² é a chamada teoria hedonista e ela é considerada uma teoria subjetivista. Mas, qual é a tese geral de uma teoria subjetivista? As teorias subjetivas a respeito do bem-estar podem assumir uma de duas teses gerais, sendo uma mais fraca que a outra; são elas:

4. Tese subjetiva fraca: o bem-estar é subjetivo no sentido que ele é constituído por estados mentais (ou desejos ou preferências);
5. Tese subjetiva forte: o bem-estar é subjetivo porque, em última instância, ele é dependente da nossa valoração. (FERKANY, 2012, p. 473)

Pensamos que (a) seja a tese aceita pela maioria dos defensores de teorias subjetivas do bem-estar. Além disso, (b) implica em uma espécie de relativismo que não parece ser aceito por defensores de teorias subjetivistas do bem-estar. Então, suponho para fins argumentativos que a tese (a) é a que melhor se aplica a todas as teorias subjetivas e é a que deve ser debatida. Dessa forma, passo a discutir a teoria de Bentham tendo em mente a tese (a).

O hedonismo foi defendido em sua forma mais clara, primeiramente por Jeremy Bentham⁴³ e, em seguida, por John Stuart Mill. Bentham é conhecido por

⁴²Para fins de exposição tratarei o termo “bem-estar” e “felicidade” como sinônimos. Todavia é possível argumentar que o conteúdo do conceito de “bem-estar” seja maior (inclua mais elementos) que o conceito de “felicidade”. Sendo possível que o segundo possa ser incluído dentro do primeiro. Todavia, esta distinção não nos auxilia aqui. Tomemos então as palavras como sinônimas. Para mais sobre o fato de que o “bem-estar” possa ter como um de seus elementos a “felicidade” além outros elementos, ver FERKANY, M. The Objectivity of WellBeing. *Pacific Philosophical Quarterly* 93, pp.472-492, 2012, principalmente, pp. 481s.

⁴³A tese hedonista não é uma criação do século XIX. Epicuro (1999) já fazia uma defesa

oferecer uma teoria hedonista mais simples que a de seu sucessor. Sua visão acerca do que seria a felicidade consistia em afirmar que os seres humanos foram colocados (pela natureza) sob dois senhores, quais sejam, o prazer e a dor (hedonismo psicológico) (BENTHAM, 1979, p.10), sendo essa a base de toda a moralidade (hedonismo ético). Assim, chamamos de hedonismo psicológico a tese segundo a qual os seres humanos são motivados pelo prazer e pela dor e chamamos de hedonismo ético a tese segundo a qual toda a moralidade deve se basear em promoção de prazer e redução da dor. Nossas ações deveriam ser julgadas com base nessa ideia simples de felicidade. Se elas (as ações) produzem ou tendam a produzir mais prazer e evitar a dor, então deveríamos executá-la. Bentham entendia “utilidade” como:

O termo utilidade designa aquela propriedade existente em qualquer coisa, propriedade em virtude da qual o objeto tende a produzir ou proporcionar benefício, vantagem, prazer, bem ou felicidade (tudo isso, no presente caso, se reduz à mesma coisa), ou (o que novamente equivale à mesma coisa) a impedir que aconteça o dano, a dor, o mal ou infelicidade para a parte cujo interesse está em pauta. (BENTHAM, 1979, p. 4).

O prazer tem seu valor mensurável através de critérios quantitativos. Segundo Bentham (1979, p.17), tais critérios seriam: (i) duração; (ii) intensidade; (iii) certeza ou incerteza; (iv) proximidade ou afastamento; (v) fecundidade; (vi) pureza e; (vii) extensão. Dessa forma, ao julgarmos uma ação, lei ou instituição, devemos buscar por aquelas que produzem mais prazer quanto às medidas de quantidade. Bentham acredita que os prazeres (e as dores) apenas se diferem um dos outros em questão de quantidade e que não há qualquer outro valor a se considerar. Então, o que determina o valor de um prazer e, portanto, de uma ação, atividade, lei etc. são nossos estados mentais positivos ou negativos, que podem ser mais ou menos duradouros, mais ou menos intensos, mais ou menos certos e assim por diante. Se isso for verdade, todos os prazeres têm igual valor, se forem iguais quanto à quantidade, ou seja, se geram o mesmo montante de

forte de, pelo menos, um tipo de hedonismo.

prazer. Se dois prazeres têm a mesma duração, intensidade, certeza etc. então, eles deveriam possuir igual valor para o indivíduo. Sobre isso, Bentham afirma que: “Preconceitos à parte, o jogo de pino é de igual valor ao das artes e das ciências da música e da poesia.” (Bentham, 1979 *apud* SINGER (org) 1994, p. 200). Essa passagem deixa claro que Bentham considerava a fonte do prazer como irrelevante para o seu valor. De fato, ele defendia que não importava quais tipos de coisas geram prazer nos indivíduos, o único fato relevante é a quantidade do prazer gerado. Porém, a despeito do que Bentham defendia acerca do valor dos prazeres (ou felicidade), nossa intuição tende a ir para o lado oposto. Tendemos a pensar que há uma diferença entre o prazer auferido de uma partida de cartas e o prazer auferido do exercício artístico, não importando a quantidade de prazer auferida do primeiro em detrimento da quantidade de prazer auferida do segundo. Essa intuição, qual seja, de que parece haver uma diferença substantiva entre os prazeres e que alguns são mais valiosos que outros, independentemente da quantidade, trouxe críticas à teoria do valor de Bentham já em sua época.

Detratores da teoria utilitarista de Bentham o acusavam de fazer uma filosofia para porcos. Thomas Carlyle⁴⁴ argumentou que Bentham estaria defendendo uma visão de vida boa digna dos animais mais simplórios da natureza, o que seria, em última instância, uma visão inadequada sobre a vida boa humana. Segundo a visão de Bentham, uma pessoa que tivesse todos os prazeres mais simples, como, por exemplo, comer, beber, ter relações sexuais etc. teria uma vida melhor do que aquele que usufrísse de prazeres que parecem mais complexos como, estudar, fazer artes, cultivar virtudes etc. desde que os prazeres mais simples ocorressem em maior quantidade. Assim, a vida de um boêmio poderia ser melhor do que a vida de um intelectual, ou, como diriam alguns interlocutores de Bentham, a vida de um porco poderia ser melhor do que a vida de uma pessoa.

Se fosse dada a escolha a um indivíduo em uma posição epistêmica favorável, tal indivíduo pensaria ser melhor a vida de um porco do que a de uma pessoa? A resposta parece ser que é preferível viver a vida humana. Ao

⁴⁴Pensador conservador escocês do século XIX que se opunha fortemente a tese utilitarista defendida por Bentham e aos pensadores que faziam parte do seu círculo utilitarista.

compararmos a vida de um porco e a vida de uma pessoa (que seja minimamente boa), tendemos a pensar que a vida da pessoa é mais desejável do que a vida do porco. Essa intuição prevalece, pois tendemos a pensar que um ser humano possui capacidades mais sofisticadas do que a capacidade de um animal e que, portanto, conseguiríamos desempenhar certas atividades que nos dariam prazeres, de algum modo, mais valiosos. Se essa intuição estiver correta, há algo de errado na teoria hedonista de Bentham. Sendo assim, a sua teoria deveria ser, pelo menos, melhorada.

É importante notar que a teoria do valor de Bentham é puramente subjetivista e puramente pessoal; uma *vida vai bem se*, e somente se, usufruirmos de uma ampla quantidade de estados mentais positivos (prazer) e temos a ocorrência de muitos poucos estados mentais negativos (dor). Não há nada que possua valor em si a não ser os estados mentais positivos. Assim, uma lei, ação, instituição etc. apenas tem valor na medida em que causa estados mentais positivos em um indivíduo ou grupo de indivíduos. Dessa maneira, a visão de felicidade defendida por Bentham é subjetivista, pois estabelece que a única coisa que faz da vida de alguém boa é a *aprazibilidade* e que o portador de “aprazibilidade” é um estado mental específico denominado “prazer”. Um ponto interessante a se notar é que, dependendo de como nós entendemos o que seja “prazer”, nosso tipo de hedonismo varia. Se entendermos que o valor do prazer advém da satisfação de nossas preferências (o prazer relevante para minha felicidade é aquele advindo da satisfação de minhas preferências), temos um tipo de hedonismo de preferência. Se por outro lado, se pensamos que o valor da experiência advém de como ela é sentida, então temos uma espécie de hedonismo fenomenológico (MULGAN, 2007). Se pensarmos que prazer deve ser entendido como uma postura positiva direcionada a algum objeto, temos o hedonismo de atitudes (*attitudinal hedonism*), como defendido por Feldman (2004). Penso que Bentham defendia uma espécie de hedonismo fenomenológico, pois ele defendia a métrica quantitativa e o igual valor entre os prazeres. É importante notar aqui que há uma diferença entre uma “noção substantiva de bem-estar” e uma “noção formal de bem-estar”⁴⁵. Tal posição

⁴⁵ Essa distinção pode ser encontrada em CRISP, R. "Well-Being", *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), Disponível em <<http://plato.stanford.edu/archives/win2014/entries/well-being/>>.

levou Bentham a negar que haveria qualquer direito fora das instituições, uma vez que direitos seriam também avaliados em função de causar estados mentais prazerosos.

Na primeira (noção substantiva), queremos responder à pergunta (a) “o que faz da vida do indivíduo, boa?”, enquanto, na segunda (noção formal), queremos responder à pergunta (b) “o que porta o valor que faz da vida indivíduo boa?”. Repare que, quando estamos falando do hedonismo, a resposta à primeira pergunta é “aprazibilidade”, enquanto a resposta para a segunda é “estado mental denominado ‘prazer’”. Explico: teorias subjetivas como o hedonismo e como a Teoria da Satisfação de Preferências (TSP) concordam na noção substantiva de bem-estar. Os defensores dessas duas teorias defendem que a resposta à pergunta (a) é “aprazibilidade”, ou seja, é a aprazibilidade (propriedade de ser prazeroso) que faz da vida de alguém melhor. Todavia elas discordam na resposta à pergunta (b). Os hedonistas defendem que o portador de aprazibilidade é um estado mental específico denominado “prazer”, enquanto defensores da TSP defendem que o portador da aprazibilidade é a satisfação das preferências. Assim, mesmo teorias do mesmo tipo (ou seja, subjetivas) podem discordar quanto ao que é substantivo e quanto à noção formal de bem-estar.

O grande problema da teoria de Bentham acerca da felicidade é que ela estabelece uma visão muito simples do que possui valor. Os critérios quantitativos de avaliação do prazer sugerem que todos os prazeres são do mesmo tipo, variando somente em quantidade em cada ocorrência. Tal posição levou Bentham a negar que haveria qualquer direito fora das instituições, uma vez que direitos seriam também avaliados em função de causar estados mentais prazerosos, então nenhum direito moral deveria existir, já que o relevante é gerar o máximo de prazer possível e não haveria qualquer demanda moral que pudesse ser baseada em um pretensão direito que existisse fora das instituições. Todavia, como discutido acima, parece haver uma diferença entre os prazeres e, tal diferença, não se resume à quantidade, intensidade etc. Mas, há diferentes tipos de prazeres? Se há, como diferenciá-los? Mill parece ter conseguido oferecer uma resposta para tais perguntas e, com isso, uma resposta à objeção de que a teoria do valor utilitarista seria “digna de porcos”. Sua teoria do valor é

tema de grande debate até hoje, pois há diferentes interpretações acerca dela⁴⁶. Ao contrário de Bentham, Mill tentou oferecer uma visão do que seria a felicidade mais robusta e que respondesse a crítica de Carlyle e seus contemporâneos. Assim sendo, passo agora a expor a teoria do valor de Mill. Primeiramente, abordo a sua visão substantiva de felicidade. Depois exponho e discuto o critério dos “juízes competentes”. Exponho rapidamente a concepção de Mill sobre os direitos morais. Por fim, apresento alguns problemas que a teoria de Mill enfrenta.

A concepção de Mill acerca do que é a felicidade pode ser colocada, pelo menos, em duas chaves interpretativas. Na primeira, felicidade é igual à presença de prazer e ausência de dor. Na segunda, felicidade é igual presença de prazer, ausência de dor, cultivo da autonomia, individualidade e virtudes⁴⁷. No que tange à primeira interpretação, seus defensores argumentam que Mill oferece claramente o conteúdo do conceito “felicidade”, quando afirma, “Por felicidade se entende o prazer, e ausência de dor; por infelicidade a dor, e a privação do prazer” (MILL, 2000, p. 187). Nessa interpretação, Mill parece oferecer uma visão claramente hedonista de felicidade, assim como fez Bentham. Já a segunda interpretação pode ser defendida como uma teoria multicomponente⁴⁸ (apesar de haver sérias dificuldades em atribuir a Mill tal interpretação). A felicidade seria composta não apenas por estados mentais, mas

⁴⁶Uma discussão sobre a interpretação da teoria do valor de Mill pode ser encontrada, por exemplo, em CRISP, 2006b, pp.19 – 62.

⁴⁷Para uma visão de que Mill aceitava as virtudes em sua concepção de vida feliz ver, por exemplo: NUSSBAUM. M. Mill entre Bentham e Aristóteles. *FUNDAMENTO – Revista de Pesquisa em Filosofia*, n. 4, jan–jun – 2012.

⁴⁸Uma teoria multicomponente é aquela que defende que o bem-estar (ou felicidade) é composto por dois ou mais elementos. Tais elementos podem ser tanto subjetivos (como estados mentais ou satisfação de preferências) quanto objetivos (como realizações, conhecimento, autonomia etc.), ou ainda ser um conjunto de elementos subjetivos e elementos objetivos. Ao passo que teorias que envolvem apenas uma componente (como, por exemplo, o hedonismo) podem ser chamadas de “teorias de componente único”. Para uma interpretação objetivista da teoria de Mill sobre o bem-estar, ver BRINK, D. ‘Mill’s Deliberative Utilitarianism’, *Philosophy and Public Affairs*, 21, 1992, pp. 67–103. A classificação entre teorias multicomponentes e de componente único é usada por Alexander F. Sarch em: SARCH. A. Multi-Component Theories of Well-Being and Their Structure. *Pacific Philosophical Quarterly* 93, pp. 439-471, 2012.

também por outros elementos que não são redutíveis a estados mentais, como, por exemplo, virtudes. Aparentemente, há uma ampla gama de interpretações acerca da visão de Mill no que tange a felicidade. Todavia, se considerarmos a primeira interpretação como a correta, Mill defende uma teoria subjetivista da felicidade, seguindo os passos de Bentham.

Supondo que Mill seguiu os passos de seu antecessor, sua tese consistiria na defesa de que os prazeres são *constitutivos* da felicidade; portanto haveria uma relação de identidade entre felicidade e prazer. Porém podemos nos questionar se a teoria de Mill também não implica uma concepção de felicidade que seja digna de porcos, como a Bentham parece implicar. A resposta é que a concepção de Mill possui uma especificidade que afasta a crítica de que concepção de felicidade é digna de porcos. Como Mill era ciente de tal crítica, ele defendeu que há uma diferença não só quantitativa entre os prazeres, como também há uma diferença *qualitativa* entre eles. Mill diferenciou dois grupos de prazer em termos de qualidade: (i) *prazeres superiores* e; (ii) *prazeres inferiores*. Todavia, qual é o critério para separarmos os prazeres do tipo (i) dos de tipo (ii)? O filósofo inglês faz essa diferenciação recorrendo ao chamado critério dos “juizes competentes”. Para sabermos quais prazeres são superiores e quais são inferiores, deveríamos recorrer ao juízo feito por indivíduos que experimentaram ao longo de suas vidas diversos tipos de prazer. Tais indivíduos, tendo experimentado uma ampla gama de prazeres, sempre hão de escolher certos prazeres em detrimento de outros. Tais prazeres que sempre seriam escolhidos pelos competentes seriam os prazeres superiores. Assim, os prazeres superiores possuiriam um maior valor em comparação com os prazeres inferiores. Mas, quais os prazeres superiores?

Os prazeres superiores estão relacionados às capacidades humanas não compartilhadas por outros animais não humanos, como, por exemplo, a inteligência, a virtuosidade, a contemplação etc. Ao passo que os prazeres inferiores estão relacionados às necessidades mais imediatas que compartilhamos com os outros animais não humanos, como, por exemplo, a necessidade de comer, beber, dormir, o desejo sexual etc. Os prazeres superiores seriam escolhidos pelos juizes, porque eles são mais *aprazíveis* do que os prazeres inferiores, não importando quão *intensos* sejam estes últimos. Uma crítica à diferenciação de prazeres pela sua qualidade é a de que ou a diferenciação é inconsistente com o hedonismo ou que Mill abandona seu

hedonismo em favor de outra teoria acerca da felicidade (como referi acima, pode haver mais de uma interpretação da teoria do valor de Mill)⁴⁹. Então, ao contrário de Bentham, Mill considerava a fonte do prazer como relevante para o seu valor, uma vez que as diferenças qualitativas dos prazeres dependem de suas fontes. Prazeres causados pelo exercício intelectual são mais valiosos do que aqueles causados por sensações corpóreas.

Com a diferenciação explicitada acima, Mill livra-se da objeção de que a teoria do valor do utilitarismo é digna de porcos, pois leva em consideração as diferenças qualitativas entre prazeres; todavia o critério dos juízes competentes não está livre de problemas. Como podemos eleger alguém como juiz competente? Deve haver sempre unanimidade entre os juízes competentes a respeito do que seja um prazer superior? O teste dos juízes pode ser circular? Se admitirmos os juízes competentes, temos de admitir que o prazer não seja o único bem, já que os juízes podem valorizar outras coisas como fim em si?

A despeito de tais dificuldades, Mill consegue oferecer uma visão mais robusta do que seja a felicidade quando comparado com seu antecessor. Sua teoria parece acomodar nossa intuição de que o prazer de ler um bom livro ou ouvir uma boa música é mais valioso (em algum sentido) do que o prazer auferido quando saciamos nossa fome. Assim, seguindo a visão de Mill, uma vida *mais* feliz seria aquela com mais prazeres superiores do que inferiores⁵⁰. Há outros problemas acerca da distinção entre prazeres superiores e inferiores. Um deles é acusação de que a distinção tem implicações elitistas/paternalistas. Repare que alguém pode dizer que a visão de Mill sobre a natureza do bem-estar implica em certo paternalismo. Mas o que isso quer dizer? Que Mill defende uma teoria que nos diz o que é bom para nós independentemente do que pensamos sobre isso. Seria como um pai dizendo a um filho o que ele (o filho) deve escolher fazer e que ele não tem outra opção a não ser seguir o que o pai diz. Isso implicaria que nossa autonomia não é relevante para a visão de Mill

⁴⁹Esta crítica pode ser encontrada, por exemplo, em BRADLEY, F.H. *Ethical Studies*, 2nd. London: Oxford University Press, 1962.

⁵⁰ Podem-se encontrar referências a essa crítica em SKORUPSKI, J. The Place of Utilitarianism in Mill's Philosophy. In: WEST, H. (org) *The Blackwell Guide to Mill's Utilitarianism*. Oxford: Blackwell Publishing, 2006, p. 57. E em, SKORUPSKI, J. *Why read Mill today?* New York: Routledge, 2006, p.33.

acerca do bem-estar. Não importa se pensamos que ter prazer através de muito estudo é ruim, no fim das contas, o que importa é que esse prazer sempre será bom, e essa característica da visão de Mill é que implicaria em um paternalismo⁵¹. Mas, essa crítica não vai muito longe, pois Mill aceita que uma vida boa pode ser composta apenas por prazeres inferiores. Então, é importante frisar que para Mill, uma vida voltada a auferir prazeres inferiores possui apenas *menos* felicidade do que aquela também voltada ao cultivo das atividades que nos permitem usufruir dos prazeres superiores. Assim, não é imoral viver como “um porco”; todavia, como seres humanos dotados de capacidades complexas, poderíamos buscar auferir prazer do uso de tais capacidades. Se for verdade que o uso de tais capacidades nos permite usufruir de prazeres qualitativamente superiores, então seria racional buscá-los, uma vez que temos interesse em promover nossa própria felicidade. Assim, seria em alguma medida irracional viver a cultivar e usufruir apenas de prazeres qualitativamente inferiores, pois nós temos as capacidades e oportunidades de ter uma vida mais feliz e valiosa do que a de um porco. No entanto, o hedonismo aceito por Bentham e Mill possui falhas na explicação de o que é o bem-estar. Vamos ver os pontos positivos e negativos deste tipo de teoria.

O hedonismo como uma teoria acerca do bem-estar possui, pelo menos, dois pontos positivos que a fazem atraente. Primeiro, (a) a teoria tem força explanatória, pois explica porque os prazeres afetam nosso bem-estar ou são constitutivos dele. A teoria hedonista defende uma visão substantiva do bem-estar, fornecendo-nos a explicação sobre o que nos torna felizes em termos de “aprazibilidade”. Ela também nos oferece uma visão formal do bem-estar, defendendo que o portador de aprazibilidade é o estado mental denominado “prazer”. Dessa forma, ela nos ofereceria uma visão completa acerca do que seja o bem-estar. Além disso, (b) a teoria explica porque algumas coisas que consideramos boas ou ruins são, de fato, boas ou ruins, conectando a explicação

⁵¹Paternalismo é a interferência, pelo Estado ou por qualquer indivíduo, na vida de um terceiro indivíduo ou comunidade sem o seu livre consentimento. Tal interferência é motivada pelo argumento de que ela trará benefícios para o indivíduo ou grupo (mesmo que o indivíduo ou o grupo não reconheçam essa interferência como provedora de qualquer benefício). Geralmente, posições paternalistas não são vistas com bons olhos, uma vez que ela implica que um indivíduo normalmente não é capaz de escolher as coisas na sua vida.

com o prazer e a dor. Todavia ela enfrenta alguns problemas que parecem ser de difícil resolução. Em linhas gerais, há dois grandes problemas: 1) nem todos os prazeres parecem ter valor em si, i.e., nem todos eles parecem ser bons para nós; 2) parece haver outras coisas além do prazer que são boas para nós.

O exemplo clássico para ilustrar o primeiro problema é o do assassino sádico. O prazer auferido por um assassino sádico em um crime brutal parece ser ruim, mesmo que o prazer seja *bom para ele*. Outro exemplo clássico é dos cristãos que eram jogados aos leões no coliseu romano. Apesar dos espectadores auferirem prazer dessa atividade, esse prazer não parece ter valor em si, pois advém de uma atividade que nos parece ruim. Intuitivamente, pensamos que esse tipo de prazer não é bom, *consideradas todas as coisas*. Porém, o hedonista não consegue afastar a ideia de que o prazer sádico é bom em si, se considerado isoladamente. O problema é: o prazer sádico é bom para a pessoa que o experimenta? A primeira reação é dar uma resposta negativa a esta pergunta. Parece haver algo de errado com a ideia de que o prazer auferido da tortura seja algo valioso não instrumentalmente para o indivíduo que o experimenta. O que nossa intuição parece nos dizer é que o prazer sádico é destituído de qualquer valor positivo, pois, se tivesse algum valor, um mundo onde o prazer sádico é usufruído seria melhor que um mundo no qual ele não exista. Aceitar que o prazer sádico tenha algum valor positivo é jogar contra nossas intuições de que nada de bom resulta de atividades sádicas. Assim, se o hedonismo não consegue afastar a conclusão de que o prazer sádico é bom, nós temos uma evidência contra tal teoria. Apesar de tal evidência não ser decisiva quanto à plausibilidade ou não da teoria, ela poderia abrir uma brecha para procurarmos por uma teoria que abarque melhor nossas intuições acerca do que seja o bem-estar.

O segundo problema é geralmente ilustrado pelo experimento mental conhecido como “máquina de experiência” de Nozick (1974, p.42-3). Esse experimento mental nos auxilia no teste para sabermos se uma teoria do bem-estar é plausível, ao defender que apenas o prazer tem valor em si. Grosso modo, o experimento funciona assim: imagine que um cientista tenha desenvolvido uma máquina capaz de simular todas as experiências que uma pessoa teria na vida real. Na máquina, a vida teria mais prazeres e muito menos sofrimento do que a vida real. Porém, ao ser ligado à máquina, o indivíduo esquece que nela está conectado. Então, é dado duas opções ao indivíduo: ou

ele continua com a sua vida normalmente no mundo real, ou é conectado a máquina de experiência e passa a ter uma vida mais agradável. A pergunta então é: o que o indivíduo a quem foi dada a escolha deve fazer?

Se o indivíduo for um hedonista ele não terá qualquer razão para *não* se conectar à máquina, visto que a melhor vida possível seria a vida conectada à máquina de experiências. Se o que faz da vida melhor são meus estados mentais positivos, então conectar-me à máquina é a escolha racional a ser feita. Todavia é contra-intuitivo pensar que uma vida conectada à uma máquina que nos estimula a ter sensações aprazíveis seja melhor que a vida não conectada à máquina, na qual temos, de fato, a chance de realizar coisas que consideramos valiosas. Assim, a vida na máquina de experiência de Nozick parece carecer de algo valioso. Nozick defende que as pessoas querem fazer de fato as coisas e não ter somente experiências das realizações (NOZICK, 1974, p. 43s). Desse ponto de vista, as pessoas não são apenas receptáculos dos estados mentais aprazíveis, elas de fato gostariam de viver uma vida com intencionalidade. Um jogador de futebol não quer apenas a experiência aprazível de ganhar a copa do mundo, ele quer ganhar de fato a copa do mundo. Então, o problema do hedonismo é defender que *apenas* os prazeres são valiosos em si e que qualquer outra coisa é um meio para auferirmos mais prazer. Como exemplifica a máquina de experiência, parece haver algo mais de valioso em nossas vidas além das experiências prazerosas. Pensamos que esse experimento mental possa ser respondido de forma plausível por hedonistas e que ele não só coloque um problema para os hedonistas como também para outras teorias do bem-estar⁵². Porém, caso a resposta dos hedonistas não seja satisfatória, nós poderíamos ligar todos à máquina de prazer, até mesmo contra a sua vontade, se o nosso objetivo é produzir o máximo de prazer possível, por exemplo. Apesar de isso não ser decisivo contra a teoria hedonista, o experimento serve para revelar nossas intuições acerca daquilo que é bom para nós. Ao menos, ele torna plausível pensarmos que pode haver algo diferente de nossos estados mentais que seja bom para nós. Esclarecido tal ponto, passemos a oferecer uma descrição da teoria da satisfação de preferências.

⁵²Uma resposta hedonista para a objeção de Nozick e para a objeção de Carlyle pode ser encontrada em CRISP (2006a, p.100-117)

2.2. Satisfação de Preferências

Uma segunda teoria desenvolvida sobre o bem-estar é a chamada teoria da satisfação de preferências (também chamada de satisfação de desejos⁵³). Em grande parte, o desenvolvimento dessa teoria se deve à crença de que uma teoria hedonista não conseguiria responder satisfatoriamente as críticas voltadas a sua concepção de bem-estar (principalmente, a máquina de experiências de Nozick). Grosso modo, esta é a tese segundo a qual o bem-estar consiste em termos de nossas preferências informadas satisfeitas. Assim, se desejamos X a Y , ter X satisfeito incrementa nosso bem-estar⁵⁴. Todavia, há vários tipos de Teorias de Satisfação de Preferências (doravante, TSP): (1) Teoria das preferências atuais: teoria segundo a qual um indivíduo I tem seu bem-estar incrementado se, e somente se, seus desejos *correntes* são realizados; (2) Teoria das preferências abrangentes: teoria segundo a qual o que importa para o bem-estar de um indivíduo I é o nível total de satisfação dos desejos em sua vida como um todo e; (2) se divide em duas vertentes: (2a) Teoria das preferências abrangentes *global*: o que importa para o bem-estar do indivíduo I é a satisfação dos seus desejos em sua vida como um todo, mas há certos desejos que possuem prioridades sobre outros e; (2b) Teoria das preferências abrangentes informadas: o que importa para o bem-estar do indivíduo I é a satisfação de seus desejos *informados* em sua vida como um todo. Abordaremos de maneira bem geral sobre cada variação, embora a que mais nos interessa é a (2b). Vamos oferecer uma visão geral de como a teoria se comporta.

De forma geral, as TSP do tipo (1) defendem que o nosso bem-estar é composto e/ou incrementado com a satisfação das preferências que temos agora (por isso desejos *correntes* satisfeitos). Isso quer dizer que nosso bem-estar apenas é incrementado se nossos desejos de agora são satisfeitos. Então, seu eu desejo beber café agora, o meu bem-estar será incrementado se eu beber o café. Não é uma ideia muito difícil de entender, não é mesmo?! Se soa plausível esta ideia, pare e pense melhor. Ao que parece, há vários desejos/preferências

⁵³Para fins de exposição, usaremos a palavra “desejo” e “preferência” como sinônimas. Para todos os efeitos, elas designam o mesmo.

⁵⁴Tais diferenciações podem ser encontradas em Mulgan (2007) e Fletcher (2016).

que temos agora que parecem não ser bons para nós, se forem satisfeitos. Por exemplo, imagine um fumante. Agora ele deseja/preferir fumar um cigarro e, se ele de fato fumar, então o bem-estar dele seria incrementado. Mas, a longo prazo a satisfação desta preferência corrente vai trazer mais mal-estar do que bem-estar. Assim, a TSP do tipo (1) não parece nos informar com precisão o que compõe/incrementa nosso bem-estar, pois se a seguirmos, podemos ter muito mais mal-estar (intuitivamente) do que bem-estar. Se esta crítica fizer sentido, nós podemos nos mover para as TSPs do tipo (2).

Teorias do tipo (2a) evitam o problema indicado acima defendendo que o bem-estar de alguém é composto pela satisfação de seus desejos/preferências como um todo, ou seja, analisada a vida como um todo. Isso quer dizer que, por exemplo, a satisfação do desejo/preferência de fumar não contribui para o bem-estar do fumante, analisada sua vida como um todo. Dessa forma abrangente, o bem-estar é entendido como a satisfação dos seus desejos/preferências de forma que as satisfações de alguns deles permitam a satisfação de outros deles. No caso do fumante, a satisfação do desejo/preferência não permitiria a satisfação de outros desejos/preferências importantes como, por exemplo, satisfazer a preferência de não morrer por um câncer de pulmão. Assim, a satisfação da preferência/desejo de não morrer tem certa *prioridade* quando comparada com a satisfação da preferência de fumar. Uma crítica que se avança para este modelo teórico é que não conseguiríamos saber se alguém está com o bem-estar alto ou baixo em certo momento da sua vida. Como o que conta é a satisfação das preferências/desejos de forma abrangente, teoricamente, nós apenas conseguiríamos dizer se a vida de alguém teve alto (ou baixo) grau de bem-estar depois que sua vida toda foi considerada. Assim, nossas avaliações de bem-estar apenas incidiriam quando a vida do indivíduo terminasse, mas se assim for, não conseguimos usar a teoria para, por exemplo, fazer avaliações morais ou mesmo avaliações econômicas (sim! Teorias econômicas utilizam teorias do bem-estar para avaliar se um sistema econômico é bom). Do ponto de vista avaliativo a teoria seria inócua, ou seja, não nos informaria adequadamente quem está com bem-estar e quem não está com bem-estar em certo momento do tempo. Sendo assim, temos razões para nos mover, novamente, entre as TSPs e aceitar uma do tipo (2b).

Em linhas gerais, a TSP (2b) é a tese segundo a qual o que compõe o bem-estar do indivíduo é a satisfação de suas preferências/desejos informados.

Dessa forma, o defensor da TSP considera que a satisfação de nossas preferências é uma consideração *necessária* e *suficiente* para o bem-estar. A diferença entre o hedonismo de preferências e a TSP (2b) é que, no primeiro caso as preferências funcionam como indicativo daquilo que trará prazer ao indivíduo, enquanto na segunda teoria as preferências *constituem* o bem-estar. Repare que, para o defensor da TSP (2b)⁵⁵, o que conta são as preferências do agente e não de terceiros, ou seja, o que conta para o *seu* bem-estar são *suas* preferências. Dessa forma, a satisfação das preferências de um sujeito S é condição necessária e suficiente para o seu bem-estar neste tipo de teoria.

A TSP tem algumas vantagens sobre o hedonismo como, por exemplo, as preferências são mais simples de serem medidas, pois elas são reveladas nas tomadas de decisão. Se for dado a escolha a um sujeito S entre as opções x, y e z e ele escolhe y, então podemos inferir que S prefere y a x e z. Esse tipo de característica fez com que a TSP fosse (e seja) amplamente utilizada em economia, como falamos acima. Em teorias como as teorias da decisão e desenvolvimentistas, a TSP é uma importante ferramenta de avaliação de estados de coisas. Outra vantagem da TSP é que ela responde melhor ao experimento da máquina de experiências de Nozick, uma vez que se conectar ou não à máquina de experiências vai ser bom para o indivíduo dependendo das preferências do sujeito. Se um sujeito S prefere viver no mundo real com uma vida com menos prazeres, então não se conectar à máquina incrementará o seu bem-estar. Por outro lado, se S prefere se conectar à máquina, então conectar-se a ela incrementa o bem-estar de S.

Algumas questões surgem ao considerarmos a TSP. Por exemplo, parece haver alguns desejos que se forem satisfeitos não trazem bem-estar nem para o indivíduo que o deseja e nem para seus pares. O problema é que a satisfação de algumas preferências não nos afeta, mesmo que ela ocorra. Imagine que o Antônio está em um voo entre o ponto São Paulo e Recife e sentado ao seu lado está o sujeito José. Durante a viagem, José conta a Antônio como a sua vida vai mal e como as coisas estão piorando para ele. Antônio, um bom sujeito, se compadece profundamente ao ver a péssima situação em que se encontra José e deseja, sinceramente, que José tenha uma mudança em sua vida e que ela melhore. Ao chegarem a Recife, Antônio pega a suas malas e se

⁵⁵Daqui em diante, vamos apenas chamar de TSP a teoria TSP (2b), para simplificarmos.

despede de José, nunca mais voltando a vê-lo. O problema é que o desejo de Antônio de ver a vida de José melhor, mesmo que satisfeito, parece não incrementar o bem-estar dele. Se ele não sabe se sua preferência foi satisfeita, não há como ela incrementar seu bem-estar.

Alguns filósofos podem desejar que todos os problemas da filosofia tenham solução, um físico pode desejar que o universo seja compreendido em sua totalidade, um educador pode desejar que todas as pessoas do mundo saibam ler e escrever, mas tais desejos são muito distantes de suas vidas. Os objetores da TSP argumentam que ela falha ao cumprir o que pode ser chamado de ‘requerimento de forte ligação’⁵⁶. O requerimento de forte ligação estabelece, grosso modo, que para uma teoria do bem-estar ser bem-sucedida na descrição do que é bom para nós não instrumentalmente, ela *não* pode aceitar que o bem-estar de alguém seja incrementado por um evento qualquer, se o evento não estiver fortemente ligado à pessoa. Dessa forma, o incremento em nosso bem-estar apenas pode ocorrer, se houver alguma mudança em nossas vidas (ou em nossos estados mentais, estado de coisas que nos cercam, ou estados relacionais). Ao que parece, uma mudança no bem-estar de uma pessoa sugere mudanças ou em seu corpo ou em sua mente e que, em alguma medida, o bem-estar deva ser experienciável. Se o defensor da TSP aceita que o bem-estar é experienciável e deve envolver algum tipo de mudança no sujeito, então ele deve rejeitar que desejos externos⁵⁷ contem para o bem-estar individual. O que chamamos de “desejos externos” são aqueles desejos que podemos ter, mas que não dizem respeito a nós mesmos, como os citados acima. Porém, como o defensor da TSP pode rejeitar os desejos externos?

Para cumprir com o “requerimento de forte ligação”, um defensor da TSP poderia restringir as preferências que contam apenas àquelas que estão fortemente ligadas ao sujeito, ou seja, apenas àquelas que impactam diretamente a vida dele. Todavia, mesmo que aceitemos tal movimento, ainda podemos perguntar: será que todas as preferências que temos sobre nós mesmos contam para o nosso bem-estar? Por exemplo, um fumante que está há

⁵⁶No próximo tópico, discutiremos mais afundo essa objeção, pois afeta também o defensor da Teoria da Lista Objetiva.

⁵⁷Para mais sobre esta crítica ver DWORKIN, R. *Levando dos Direitos a Sério*. São Paulo: Martins Fontes, 2002, pp. 357-367.

cinco dias sem fumar, por estar tentando parar por causa de um problema de saúde, deseja fumar um cigarro; então, para cumprir com a sua preferência, nós lhe damos um cigarro. Após fumar o primeiro cigarro, o fumante volta a fumar como antes e seu problema de saúde se agrava. Por mais que o fumante tenha satisfeito a sua preferência, isto não parece ter incrementado seu bem-estar. Então, ao forçar a teoria a cumprir o “requerimento de forte ligação”, o defensor da TSP cai em outro problema, qual seja, nem todas as preferências que estão ligadas fortemente ao indivíduo, se satisfeitas, parecem incrementar seu bem-estar.

Para responder a esse problema adicional, o defensor da TSP recorre, geralmente, a duas estratégias. A primeira é a estratégia de exigência de informação completa. A segunda é a exigência de prudência. Assim, as preferências que contam para o nosso bem-estar são aquelas que formaríamos quando bem informados e de modo prudente. Se considerarmos essas duas exigências, conseguimos responder ao problema das preferências que são prejudiciais. De um ponto de vista epistemicamente favorecido, o fumante prudente não gostaria de fumar aquele primeiro cigarro que o levará novamente ao vício, pois o que ele de fato desejaria é a saúde plena e não a doença. Apesar de a estratégia ter algum sucesso em afastar as preferências que são prejudiciais, ela ainda incorre em risco de circularidade. Como diz Mulgan (2012, p.107):

O desafio para qualquer um desses testes é evitar a circularidade (Pergunta: Quais desejos melhoram o bem-estar? Resposta: Aqueles que uma pessoa prudente plenamente informada teria. Pergunta: Quais desejos uma pessoa prudente e plenamente informada teria? Resposta: Aqueles que melhoram o bem-estar.)

Mesmo que o defensor da TSP não caia na circularidade, ainda pode haver problemas com preferências que não são prejudiciais, mas que são sem sentido. Imagine que João tem o desejo de escrever todos os números de zero a um milhão. Se compararmos esse desejo com o desejo de, por exemplo, ensinar boa filosofia a alunos de graduação, nós veremos que o primeiro parece menos valioso que o segundo. Mas por que parece que o primeiro vale menos que o

segundo (parece que não incrementa nosso bem-estar como o segundo)? Se pensarmos que o primeiro de fato vale menos que o segundo, então nossa intuição nos diz que há algo mais valioso em “ensinar boa filosofia” que independe da preferência, pois se assim não fosse, o desejo de escrever todos os números de zero a um milhão e de ensinar filosofia teriam o mesmo valor.

Dado tais problemas deveríamos buscar alguma alternativa à teoria da satisfação de preferências, por ela parecer não nos oferecer uma visão plausível acerca do bem-estar. O fato de nem toda satisfação de preferência ser *boa* para o indivíduo já aponta para algo de errado em uma teoria que pretende estabelecer o que é *bom* para os indivíduos⁵⁸. Porém, a TSP sofre em larga medida do problema mais geral que afeta várias formas de hedonismo. Não dar a devida ênfase aos estados de coisas e apenas enfatizar a satisfação das preferências coloca os defensores da TSP a enfrentar problemas parecidos com aqueles de algumas formas de hedonismo. Ao desconsiderar ou não dar a devida relevância aos estados de coisas do mundo, para a explicação do que é valioso não instrumentalmente para o indivíduo, a TSP tem de lidar com as distorções causadas na explicação daquilo que é o bem-estar. Podemos notar isso quando discutimos preferências sem sentido ou preferências externas: apesar de as satisfazermos, elas não têm qualquer efeito sobre nosso bem-estar, já que os estados de coisas não são devidamente considerados - no primeiro caso, os estados de coisas (as situações a qual se refere às preferências) não possuem qualquer valor, enquanto que no segundo caso, os estados de coisas estão completamente desvinculados da vida do indivíduo. Pensar nos estados de coisas como relevantes para explicação do que seja bom não instrumentalmente para um indivíduo, nos coloca a caminho de uma teoria objetivista.

2.3. Teoria da Lista Objetiva

Uma terceira teoria acerca do bem-estar é a Teoria da Lista Objetiva (TLO). Essa teoria defende, em linhas gerais, que há mais de um tipo de estado

⁵⁸ Uma discussão mais detalhada da teoria da satisfação de preferências e seus problemas pode ser encontrada em ARNESON, R. Human Flourishing Versus Desire Satisfaction. *Social Philosophy & Policy Foundation*. 1999, p. 113- 142, especialmente, das páginas 126 a 135.

de coisas (ou estados relacionais) que incrementam nosso bem-estar e que *algumas* dessas coisas independem de nós as preferirmos ou auferirmos estados mentais positivos (prazer) ao usufruirmos delas. Tal teoria oferece uma lista de bens gerais que incrementam ou compõem nosso bem-estar, e a lista pode conter elementos tanto objetivos quanto subjetivos. Alguns desses bens gerais incrementam nosso bem-estar independentemente das nossas preferências para com eles. Outros incrementam nosso bem-estar apenas enquanto temos uma atitude positiva acerca deles, por exemplo, o conhecimento pode incrementar nosso bem-estar somente se nós buscarmos adquiri-lo. A lista contém, via de regra, bens gerais que fazem a vida de uma pessoa ir melhor. Portanto a lista contém vários bens gerais que incrementam ou são constituintes do bem-estar do indivíduo, se eles são instanciados em eventos e reconhecidos pelos indivíduos. Tais elementos, argumenta-se, possuem uma “propriedade fazedora-de-bondade”. Isso significa que os elementos que compõem o bem-estar têm alguma característica que fazem deles bons por si mesmos e, portanto, nos beneficiam não instrumentalmente.

Os tipos de elementos tomados de forma mais abstrata serão chamados de ‘bem gerais’. Um ‘bem geral’ é o nome dado a um conjunto de coisas que podem ser unidas sob o mesmo nome e que são boas. Por exemplo, “ser saudável” possui valor intrínseco de forma geral, mas há várias maneiras de ser saudável e, cada uma dessas maneiras, também são boas. Assim, um bem geral é um conjunto de coisas que possuem similaridades como, por exemplo, a mesma propriedade fazedora-de-bondade. Bens desse tipo são bens universais, ou seja, eles são bons para todos os indivíduos dada propriedade universal do bem geral. Já um bem a nível concreto, é algo que é bom para um indivíduo se, e somente se, ele é um exemplo de um bem geral. Assim, um evento, ou estado de coisas, etc. é bom em si mesmo se, e somente se, ele instancia uma propriedade fazedora-de-bondade de algum bem geral. Tomem um exemplo, assume-se que relacionamentos afetivos sejam um bem geral, ou seja, um conjunto de instâncias que compartilham a mesma propriedade fazedora-de-bondade. Vamos assumir, para fins de argumentação, que tal propriedade seja “amor recíproco”. Eventos particulares que instanciam o bem geral “relacionamentos afetivos” podem ser, por exemplo, “amizade entre A e B”, “relação parental entre A e B” etc. Assim, um indivíduo pode ser beneficiado pela ocorrência de um evento particular se o evento é uma instância (um exemplo, um

modo de ser) de um bem geral. Um defensor da teoria da lista objetiva deve responder algumas perguntas centrais ao compor a sua lista, como:

- (a) O que deve compor a lista?;
- (b) Qual o “fazer-de-bondade”?;
- (c) Como decidimos o que compõe a lista? (CRISP, 2006a, p.101-102).

A pergunta (a) questiona quais são os elementos que podem entrar em uma lista. A resposta mais simples para ela é que o que deve entrar na lista são aquelas coisas que fazem da nossa vida melhor. Bem, a resposta é simples, mas pouco (ou nada) informativa. Há, pelo menos, duas perspectivas sobre quais bens gerais devem compor a lista. A primeira é seguir a resposta simples e apenas oferecer elementos para lista que, intuitivamente, nos parecem importantes para que uma pessoa obtenha bem-estar. A segunda é tentar identificar as propriedades fazedoras-de-bondade dos elementos que devem compor a lista⁵⁹. Sobre a primeira perspectiva, Crisp (2006b, p. 102) afirma que teorias que apenas oferecem uma lista dos elementos do nosso bem-estar são teorias *enumerativas* do bem-estar. Uma TLO que apenas apresenta a lista de elementos que compõem o bem-estar sem dar-lhes uma boa justificação é uma teoria *enumerativa* da lista objetiva. Repare que todas as teorias do bem-estar nas quais não é apresentada uma visão substancial do porquê algum elemento compõe o bem-estar torna-se uma teoria *enumerativa* (uma teoria da preferência pode ser enumerativa, por exemplo).

Teorias *enumerativas* do bem-estar não oferecem resposta à pergunta (b). A pergunta (b) versa sobre o que torna algo bom. Se defendermos que o prazer é bom, podemos legitimamente perguntar: mas o que torna o prazer bom? A resposta é uma visão substantiva, o que torna o prazer bom é a sua “aprazibilidade”. A *maioria* dos defensores da TLO não oferece (ou considera

⁵⁹ Ao tentarmos responder (a), podemos ter que nos comprometer a responder (b) primeiro. Por exemplo, podemos defender que os elementos que compõem a lista que descreve o bem-estar são aqueles que possuem uma propriedade fazedora-de-bondade X. Porém não é necessário responder (b) para responder (a). Podemos apenas assumir uma teoria enumerativa do bem-estar e termos uma resposta à (a), mas, ao fazermos isso, assumimos que não há resposta a (b). Uma teoria substantiva daria uma resposta a (b) e através dela uma resposta a (a).

impossível oferecer) uma resposta a (b). Porém, há um tipo de TLO que consegue oferecer uma resposta para (b), qual seja a TLO perfeccionista. A título de esclarecimento, Lauinger (2013) divide as TLO em dois grupos, quais sejam, (i) Teoria da Lista Objetiva de lista-bruta e, (ii) Teoria da Lista Objetiva perfeccionista. Segundo ele (2013, p.955)

O teórico da lista objetiva de lista-bruta considera os bens gerais da lista sendo bens prudenciais e não bens perfeccionistas, ao passo que o teórico da lista objetiva perfeccionista considera todos os bens gerais na lista não somente como bens prudenciais, mas também como bens perfeccionistas.

O importante, agora, é notar que o defensor da TLO perfeccionista é capaz de oferecer uma resposta para a pergunta (b). Para um perfeccionista, a propriedade “fazedora-de-bondade”, ou seja, a característica que faz dele bom por si mesmo é a sua contribuição para o aperfeiçoamento da natureza humana. Assim, algo é constitutivo do nosso bem-estar se, e somente se, contribui para que alcancemos o aperfeiçoamento da natureza humana. Portanto, a tese perfeccionista defende que deve haver uma espécie de adequação entre os elementos do nosso bem-estar e nossa natureza. Como o defensor da TLO perfeccionista pode oferecer uma resposta à pergunta (b), ela não é apenas uma teoria *enumerativa*, mas sim, uma teoria *substantiva* ou *explanatória*.

Teorias *substantivas/explanatórias* do bem-estar são capazes de justificar porque os elementos que compõem o bem-estar são portadores de uma propriedade específica que faz deles “bons”. É importante notar que as noções de teoria enumerativa e de teoria substantiva, não são excludentes. Dessa forma, uma teoria que seja enumerativa também pode ser substantiva/explanatória⁶⁰. A resposta à pergunta (a) de um perfeccionista é que os bens que vão à lista são aqueles que aperfeiçoam a natureza humana. A resposta à pergunta (b) é que a propriedade fazedora-de-bondade de algo é a sua contribuição para o

⁶⁰Para uma defesa de que teorias enumerativas e teorias substantivas/explanatória podem coexistir, ver: FLETCHER, G. A Fresh Start for the Objective-List Theory of Well-Being. *Utilitas*, Vol.25, p.206-220, principalmente: pp. 208-209.

aperfeiçoamento da natureza humana. E a resposta à pergunta (c) seria que teríamos que identificar a propriedade fazedora-de-bondade nos elementos que estamos a considerar para incluí-los (ou não) na lista. Acerca de (c), o defensor não perfeccionista da TLO poderia eleger o que vai à lista através de julgamento reflexivo ou intuição, ou seja, escolher os elementos da lista através da observação daquilo que parece compor o bem-estar.

Isso posto, podemos nos perguntar: afinal, quais são os bens gerais que compõem a lista? Há uma discordância entre os proponentes da TLO sobre quais elementos devem compor a lista. Todavia, geralmente, uma lista é composta dos seguintes elementos:

- Realizações (Fletcher, 2013, p. 214; Rice, 2013, p.200)
- Conhecimento (Rice, 2013, p.200);
- Autonomia, Liberdade, capacidade de agir (Varelius, 2003; Ferkany, 2012, p.481; Rice, 2013, p.200);
- Prazer (Mill, 2000, p.187; Fletcher, 2013, p. 214; Ferkany, 2012.p.481; Rice, 2013, p.200);
- Satisfação de Necessidades básicas (Griffin, 1986, p.40);
- Amizade (Fletcher, 2013, p. 214, Ferkany, 2012, p.482, Griffin, 1986, p.67, Parfit, 1984, p. 314);
- Virtudes (Hurka, 1993, p.9-17);
- Respeito e autorrespeito⁶¹;

Uma lista deste tipo pode incluir outros elementos, dependendo da abordagem que empregamos. Alguns outros elementos que podem constituir a lista são matéria de debate. Alguns dos itens são mais controversos que outros.

A TLO possui algumas qualidades que suas concorrentes (Hedonismo e TSP) não possuem. Por exemplo, a TLO é imune à objeção da “máquina de experiência” uma vez que o que é bom para os indivíduos não depende somente dos estados mentais ou seus desejos. Para o defensor da lista objetiva, o valioso é realizar *de fato* algo e não ter *somente* a *sensação* de que realizou algo. A

⁶¹Adaptado a partir de MULGAN 2012, p.120.

teoria também pode afastar os prazeres que nos parecem ruins, pois, *todas as coisas consideradas*, o prazer de um assassino não incrementa o bem-estar. Se imaginarmos uma vida considerada como um todo com o prazer sádico e uma vida sem prazer sádico, provavelmente, consideraríamos a segunda melhor que a primeira. A TLO também sana a nossa intuição de que há coisas que são valiosas independentemente de auferirmos prazer dela ou de a desejarmos. Para a TLO ter prazer ou ter uma preferência satisfeita não é uma condição suficiente para ter bem-estar.

Ao contrário do hedonismo e da TSP, a TLO *pode* defender que ter prazer e satisfação de preferências é uma condição necessária, mas não suficiente, para o bem-estar⁶². Outro ponto positivo é que a TLO nos oferece uma base informacional mais acessível para avaliarmos os estados de coisas. Tomando uma tese hedonista, podemos encontrar dificuldades para fazer comparações interpessoais de bem-estar. Para o hedonismo, o que conta como elemento constitutivo do bem-estar são os estados mentais que são acessíveis apenas em primeira pessoa. Vimos que a TSP pode nos dar melhores informações para a comparação, já que as preferências são reveladas nas escolhas dos sujeitos. A TLO pode ser ainda melhor nesse quesito, pois parte da informação que é necessária para fazer as comparações interpessoais. Graças a essas vantagens a TLO parece ser uma teoria do bem-estar mais satisfatória.

Por outro lado, a TLO também sofre de problemas. O primeiro problema é a variedade entre as listas. Defensores da TLO discordam sobre quais elementos devem fazer parte da lista e quais devem ficar de fora. Parece não haver um critério único ou um método claro que nos permita incluir ou excluir bens gerais da lista. Repare que o problema não é a discordância dos teóricos (discordância entre especialistas existe em todas as áreas do conhecimento), mas sim a falta de um critério que nos permita incluir e excluir elementos da lista - pelo menos, esse critério é inexistente nas TLO de lista-bruta. A falta de um critério claro que nos possa auxiliar a escolher quais bens fazem parte da lista *pode* tornar qualquer lista arbitrária. A falta do critério também nos deixa em dúvida acerca de se uma lista é correta e exhaustiva, ou seja, que a lista capta o que é o bem-estar. Outro problema é que se construímos uma lista que se pretende universal, ela pode ser insensível às diferenças culturais, não deixando

⁶²Ver, por exemplo, ADAMS 1999.

espaço para o pluralismo cultural. Uma segunda objeção a se considerar é a de que algum item da lista possa ser ruim para alguém. Outra falha da TLO pode estar em supor que todos os bens gerais que vão à lista são bons para todos os indivíduos, independentemente dos eventos e contextos de instanciação. Esses são problemas que qualquer teórico da TLO tem que enfrentar.

Em resumo, a TLO estabelece, grosso modo, que o que é bom para nós não instrumentalmente é composto por um conjunto de elementos que tem valor independentemente de nossos estados mentais ou preferências. Tais elementos são bons, mesmo que não o reconheçamos seu valor, ou mesmo que não tenhamos nenhum estado mental positivo causado por eles. Assim, esse tipo de teoria defende que o que é importante para fazer nossa vida melhor não é apenas ter certo tipo de estados mentais. As coisas importantes que fazem de nossa vida melhor são uma variedade de estado de coisas, eventos etc. que nós podemos acessar. Assim sendo, a principal diferença entre teorias subjetivas e a TLO é que a última defende que o nosso bem-estar não é completamente dependente de nossos estados mentais ou desejos. Para ter bem-estar (ou incrementá-lo), nós precisamos ter acesso às instâncias de bens que são boas em si mesmas. Consequentemente, o que é bom para nós não instrumentalmente depende também de como é o mundo. O problema central para o proponente da TLO é explicar *por que* os estados de coisas, eventos ou estados relacionais são bons por si mesmos e *como* eles podem nos beneficiar diretamente.

Apesar de termos estas várias teorias do bem-estar disponível, ainda é difícil bater o martelo e dizer qual delas é a melhor. O modelo hedonista foi o mais aceito por muito tempo, por outro lado, a TSP é a mais influente em várias áreas hoje em dia (em especial, na economia) e o modelo da TLO está em profundo desenvolvimento, atualmente. Nesta disputa acirrada entre modelos de explicação sobre o que é o bem-estar, todos eles estão virtualmente empatados.

3. Metaética

Como dissemos em nosso ponto 1, a metaética estuda os aspectos mais gerais e abstratos da Ética. Geralmente, as teorias em metaética são divididas, recorrendo-se ao tipo de respostas que elas dão aos problemas epistemológicos ou metafísicos da ética. Em metaética, quando estamos

preocupados com problemas metafísicos, queremos responder problemas como: “será que as propriedades ou fatos morais existem? Se eles existirem, qual a natureza dos mesmos? Serão eles iguais aos fatos naturais, ou serão eles *sui generis*? E se não houver fatos morais, quão problemático isso é para a Ética? Afinal, se não houver fatos morais, a respeito do que a Ética fala? Nós temos três modelos mais gerais para responder essas questões: (i) o Realismo; (ii) o Quase-realismo e; (iii) o Irrealismo (ou antirrealismo) – veremos com mais detalhes abaixo⁶³.

Já quando estamos lidando com problemas metaéticos epistemológicos, estamos preocupados em responder questões como: será que há conhecimento moral? Como seria possível estabelecer o valor de verdade dos juízos morais? Quais as justificações que seriam aceitas para as nossas reivindicações de conhecimento moral? Acerca desses problemas, nós temos dois modelos teóricos mais gerais: (I) O Cognitivismo, e (II) O Não-Cognitivismo. Assim, podemos ter várias combinações entre as respostas dos dois conjuntos de problemas – reparem que os problemas metafísicos e epistemológicos são logicamente independentes, de modo que há uma série de combinações entre realismo, quase-realismo e irrealismo com o cognitivismo e não-cognitivismo. Tomando as teorias cognitivistas e não-cognitivistas como principais, temos as seguintes teorias disponíveis: (a) Cognitivismo Naturalista (realismo naturalista): (a1) Cognitivismo Naturalista Reducionista e, (a2) Cognitivismo Naturalista Não-Reducionista; (b) Cognitivismo Não-Naturalista (realismo platônico); (c) Cognitivismo não-realista (qualquer tipo): (c1) A teoria do Erro de Mackie; (c2) Teorias dependentes-de-respostas; (d) Teorias não-cognitivistas⁶⁴. Vamos começar com o Cognitivismo Realista (primeiro o naturalista e depois o não naturalista). Depois comentaremos brevemente a teoria do erro de Mackie. Em um terceiro momento falaremos do não-cognitivismo.

⁶³Nossa exposição neste subcapítulo segue a nomenclatura e estrutura estabelecidas por Alexander Miller, 2013.

⁶⁴ Classificação baseada na feita por Alexander Miller em *Contemporary Metaethics An Introduction*. Cambridge: Polity Books, 2013.

3.1. Cognitivismo Realista e sem Realismo

Teóricos que defendem a posição cognitivista são aqueles que respondem perguntas como “há conhecimento moral?” de modo positivo (mas, como veremos, pode-se defender que a moralidade deveria gerar conhecimento, mas não o faz, por exemplo, por não haver verdades morais para conhecermos, dada a inexistência de fatos morais). *Geralmente*, quem é cognitivista assume também alguma posição realista, pois responde de forma positiva à pergunta “há fatos morais?”. A discordância mais profunda entre cognitivistas realistas é sobre as questões metafísicas, ou seja, qual a natureza dos fatos morais. Assim, comecemos com o Cognitivismo Naturalista. Grosso modo, o cognitivismo naturalista é a tese segundo a qual há conhecimento moral e que, quando fazemos afirmações acerca da moralidade, estamos apontando para fatos naturais que fazem das nossas afirmações verdadeiras ou falsas – ou como se diz “as proposições morais possuiriam valor de verdade”. Assim, cognitivistas naturalistas consideram que a moralidade é expressa em proposições de fato e que os fatos morais são de alguma forma relacionados aos fatos naturais. Relacionados como? Neste ponto temos dois tipos de naturalismo: o reducionista e o não reducionista. Para os primeiros, as propriedades morais **são** redutíveis aos fatos naturais. Então, de algum modo, eles negam a distinção fato/valor⁶⁵ ou entendem a distinção de modo a não cometer a falácia naturalista. Por exemplo, reducionistas podem defender que a propriedade de “ser bom” pode ser definida em termos de um estado mental (natural) denominado “prazer” – isso que seria a “redução”: definir um termo moral em termos não morais. Um naturalista reducionista famoso é o filósofo Peter Railton. Ele defende que

[...]as propriedades morais podem ser reduzidas às propriedades não morais, mas tais propriedades não são propriedades físicas no mesmo sentido que são as propriedades de massa ou velocidade, por exemplo. As propriedades não morais às quais são reduzidas

⁶⁵Acerca da distinção fato/valor ver CRISP, R. A distinção fato/valor. Disponível em https://criticanarede.com/eti_factovalor.html#:~:text=Um%20facto%20%C3%A9%20um%20estado,que%20o%20prazer%20%C3%A9%20bom,2006c.

as propriedades morais são constituídas de “interesses objetivos” de um agente epistêmico ideal ou de um “espectador imparcial”. Assim, o realismo de valor de Railton apela para um agente em uma situação epistêmica privilegiada com a finalidade de estabelecer um interesse subjetivo *objetivado* que indica o que é bom não-moralmente. Um agente A+ em posição epistêmica favorável seria capaz de identificar o interesse subjetivo objetivado de um agente A que estivesse em uma situação epistêmica desfavorável.” (SANTOS & MERLUSSI, 2013, P.349)

Já o segundo grupo de naturalistas defende que os fatos morais não são redutíveis aos fatos naturais. Apesar de estarem relacionadas, propriedades morais e propriedades naturais são diferentes, mas podem ser observadas indiretamente, assim como algumas propriedades físicas, por exemplo.

Além dos cognitivistas naturalistas (reducionistas ou não-reducionistas), ainda temos os cognitivistas não-naturalistas. O nome clássico que defendeu este tipo de visão foi G.E. Moore em seu livro *Principia Ethica*. Nele, dentre outras coisas, Moore defende que as propriedades avaliativas (ser bom/mal, correto/incorreto etc.) não podem ser nem reduzidas, nem estar relacionadas às propriedades naturais, apesar de serem existentes. Moore defendeu que tais propriedades são *sui generis*, ou seja, elas são existentes, mas não da mesma maneira que são as propriedades físicas. Moore invocava, portanto, uma espécie de realismo platônico acerca das propriedades e fatos morais. Em última instância, segundo Moore, os conceitos morais são não analisáveis em termos não morais; são, ao invés, primitivos. Isso quer dizer que eles não podem ser explicados em termos de outros conceitos – isso que significa, não ser analisável. Mas como Moore chega a esta conclusão? Para entendermos sua posição temos que entender o que ele pensava sobre o que é uma análise conceitual bem-sucedida. Segundo Moore, uma análise conceitual bem-sucedida deveria satisfazer três critérios: (a) sinonímia; (b) verificabilidade e; (c) extensionalidade. O critério mais relevante para nós é o de sinonímia. Ele pode ser expresso como se segue: qualquer expressão que represente o conceito a ser analisado é

sinônima a qualquer expressão que represente o conceito ao qual o conceito a ser analisado é reduzido (SANTOS & MERLUSSI, 2013). Assim, a análise “X é solteiro se, e somente se, é não casado” satisfaz o critério de sinonímia. Mas o que isso tem a ver com a análise dos conceitos morais em termos não morais? Segundo Moore, nenhuma análise de um conceito moral em termos não morais satisfaz a sinonímia. Por exemplo, imagine que queremos analisar o conceito de “bom” em termos de “prazer”, então definimos que “algo é bom se, e somente se, algo é prazer”. Reparem que os dois lados da análise (separados pelo conectivo lógico “se, e somente se”) não são sinônimos, então, segundo Moore, esta não é uma análise bem-sucedida. Para testar isso, nós poderíamos perguntar: “será que prazer é bom?”. Essa pergunta faz sentido (ou seja, parece ser uma pergunta legítima). Mas se ela faz sentido então quer dizer que a nossa definição falha. Pois, segundo o critério de sinonímia, a pergunta deveria não fazer sentido (ou seja, não ser legítima); seria como perguntar algo como “será que prazer é prazer?”. Este argumento de Moore ficou conhecido como “argumento da questão em aberto”. Se a pergunta que fazemos demanda uma resposta substantiva (está aberta a diferentes respostas), então nossa análise não obteve sucesso e, portanto, não é possível definir um conceito moral através de um não moral. Por outro lado, se a nossa pergunta não assume respostas além da constatação do significado dos conceitos que encerra, então nossa análise foi bem-sucedida, já que respeita o critério de sinonímia⁶⁶. Se é impossível reduzir propriedades morais a não morais, então o estatuto ontológico delas deve ser *sui generis*. Sendo assim, Moore defendia que a Ética exprime proposições, ou seja, que os juízos morais podem ser verdadeiros ou falsos em virtude de fatos morais de natureza platônica⁶⁷.

Adicionalmente, Moore acusava aqueles que pretendiam fazer uma análise de conceitos morais em termos não morais de cometerem uma falácia, a chamada falácia naturalista:

⁶⁶Para ver que a concepção de análise bem sucedida de Moore é falha, consultar, por exemplo, HARTMAN, R.S. The Definition of Good: Moore's Axiomatic of the Science of Ethics. In *Proceedings of the Aristotelian Society New Series*, Vol. 65 (1964 - 1965), pp. 235-256.

⁶⁷Um defensor contemporâneo deste tipo de posição é McDowell. MCDOWELL, J. *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, 1998.

Moore iniciou uma controvérsia em filosofia moral ao afirmar que, ao menos até ele, a filosofia moral fora deturpada por uma falácia, a falácia de analisar o conceito de bom por meio de conceitos não-morais. Em outras palavras, Moore defendeu que o conceito de bom é simples e não-analisável; não sendo analisável, em particular, em termos não-morais. O célebre termo “falácia naturalista” foi cunhado pelo filósofo precisamente para rotular as perspectivas filosóficas que assumiam que o conceito de bom poderia ser analisável por meio de conceitos não-morais. Na perspectiva de Moore, a falácia naturalista estava presente, por exemplo, nos trabalhos de Jeremy Bentham, Herbert Spencer e T. H. Green. O argumento que o filósofo apresentou a favor de sua afirmação ficou conhecido na literatura como o argumento da questão em aberto” (SANTOS & MERLUSSI, 2013, P.347)

Dado a falha na aplicação do argumento de Moore da questão em aberto, contemporaneamente, muitos filósofos discordam que exista qualquer falácia na tentativa de reduzir propriedades morais às propriedades não morais. Segundo Railton, Darwall e Gibbard (2013, p.15)

Há já cinquenta anos no entanto que sabemos que Moore não descobriu falácia alguma. Além disso a aplicação desastrosa do seu famoso “argumento da questão aberta”, feita em defesa de suas afirmações, apelava para um agora já defunto Platonismo intuicionista, que envolvia ainda certas suposições a respeito da transparência dos conceitos e da evidência da verdade analítica, as quais levavam inescapavelmente ao “paradoxo da análise” (algo que, no final das contas, era visto por Moore ele mesmo (sic)).

Todavia ainda parece ser um problema em aberto saber se reduzir as propriedades morais às propriedades não morais é um passo legítimo em Metaética. Aqui, vale alertar o leitor que não confunda a suposta falácia naturalista com o problema *é/deve*. A dita falácia naturalista é um problema metaético, no qual tentamos analisar uma propriedade moral (bondade, correteza etc.) através de uma propriedade natural (Moore pensava que isso era um erro de raciocínio). Já o problema *é/deve* (*is/ought* ou *Lei de Hume*) é quando passamos de descrições para prescrições. Por exemplo, do fato de que *S está* passando fome (descrição de estado-de-coisas), não se segue que *S deve* ajudá-lo a não ter fome (prescrição/normatização). Reparem que os dois enunciados têm naturezas distintas e um não pode ser extraído logicamente do outro. O que está por detrás do problema *é/deve* (ou Lei de Hume), segundo Crisp (2006c, p.1) é

A dita “Lei de Hume” se entendermos que é acerca de avaliações, sugere que estas não são tentativas puras de exprimir factos. Uma versão famosa e influente desta perspectiva é a de Hume, que afirmou que as conclusões com “deve” não se seguem logicamente de afirmações com “é”. Logo, se afirmamos correctamente que algo deve ser feito (e isso pode ser um dos nossos valores) com base num argumento que aparentemente se refere apenas a factos, uma das afirmações “factuais” envolve um “deve” oculto.

Assim, devemos ficar atentos para não misturar os dois problemas; ademais, é importante questionarmos se eles são de fato um problema para cognitivistas naturalistas. Afinal, a chamada Lei de Hume se assenta em uma distinção problemática entre “fatos” e “valores”.

Ressalva feita, continuemos com a nossa exposição acerca das teses cognitivistas. Agora vamos abordar a tese de Mackie. Dentro do cognitivismo temos a estranha teoria do erro de John Mackie. Por que ela é estranha? Porque Mackie defendeu que os juízos morais possuem valor de verdade (o que faz dele um cognitivista), mas que todos eles são falsos, dado que não existiria qualquer

coisa como fatos ou propriedades morais (o que faz dele um irrealista/não-realista). Ainda segundo Mackie, se defendermos que tais coisas existem, seríamos defensores de uma metafísica completamente estranha, como, por exemplo, ter que defender que os fatos/propriedades morais são existentes diferentes dos fatos/propriedades naturais, ou que por mais que nós não consigamos constatar nenhum fato/propriedade moral, eles estariam em algum lugar no mundo externo. Thomas Nagel em sua Tanner Lecture “*The Limits of Objectivity*” em 1970 (p.114) explicou a posição de Mackie como se segue:

Em uma discussão, Mackie afirmou que eu o havia interpretado mal e que sua descrença na realidade dos valores e razões não depende da suposição de que para serem reais eles devem ser entidades ou propriedades estranhas. Como ele diz em seu livro, isso se aplica diretamente às próprias razões. Pois sejam lá o que elas forem [ou seja, independentemente de se tais propriedades forem estranhas ou não], não são necessárias para explicar nada do que acontece e, conseqüentemente, não há razão para acreditar em sua existência.

Em última instância, Mackie oferece como uma razão para aceitarmos sua posição que caso houvesse propriedades morais, sua mera constatação nos motivaria a agir. O problema é que isto não ocorre e, portanto, temos mais razões para pensar que tais coisas não existem do que razões para defender o contrário⁶⁸. Sendo assim, nosso pensamento moral nos levaria a um erro, qual seja, sustentar que existem coisas como fatos/propriedades morais e, por isso, nossos juízos morais poderiam ser verdadeiros.⁶⁹ A exposição aqui feita não esgota todas as teorias disponíveis. Mas, a classificação exposta nos dá uma

⁶⁸Ao que parece, Mackie estaria apenas aplicando a Navalha de Ockham e oferecendo a melhor explicação da realidade. Neste caso, a explicação mais econômica em termos de entidades.

⁶⁹Para mais da Teoria do Erro de Mackie, ver MACKIE, J. *Ethics: Inventing Right and Wrong*. New York: Penguin, 1977.

ideia bastante clara de quais teorias são possíveis. Todavia ainda não falamos dos não-cognitivistas. Vamos à exposição geral deste tipo de teorias.

3.2. Não-cognitivistas não-realistas e quase-realistas

Não-cognitivistas sustentam, de forma geral, uma tese epistemológica muito forte, qual seja, a de que não há qualquer ganho de conhecimento quando fazemos juízos morais. Assim sendo, a moralidade não pode ser expressa em proposições, ou seja, nossos juízos morais não possuem valor de verdade. Estrito senso, a moralidade, segundo não-cognitivistas, expressa estados psicológicos de certo tipo, portanto tais expressões não são nem verdadeiros nem falsas. Em outros termos, dizer algo como “*Eu desejo X*” tem a sua verdade dependente da existência de um estado psicológico em quem profere a frase. Desejos podem ser expressos por pessoas distintas, o valor de verdade da expressão de um desejo depende exclusivamente de quem o expressa. Dessa forma, qualquer expressão como “eu desejo X” será verdadeira quando expressa por Maria que possui tal desejo e falsa para Joana que não o possui. O problema é justamente a variância da verdade, ou seja, dado que as mesmas expressões de desejos podem variar em valor de verdade dependendo de quem fala, então elas não podem ser portadoras de valor de verdade genuínas, dado que a verdade de uma proposição, por exemplo, é invariável e independente de quem fala. Dessa forma, se os juízos morais expressam apenas estados psicológicos como sentimentos, atitudes, disposições etc. eles não portam a verdade e, portanto, não há ganho cognitivo significativo quando os expressamos (MILLER, 2013). Adicionalmente, não-cognitivistas não parecem aceitar a existência de fatos/propriedades morais (mas essa não é uma condição necessária para ser não-cognitivista).

O principal argumento não-cognitivista está baseado no que chamamos de psicologia moral. Miller (2013, p.6) expressa esse argumento da seguinte forma:

Suponha que os juízos morais podem expressar crenças, tal como o cognitivista argumenta. Estar motivado a fazer algo ou perseguir um curso de ação é sempre uma questão de ter uma crença e um desejo. Por exemplo, estou motivado a ir até a geladeira porque acredito que

ela contém uma cerveja e eu tenho um desejo por uma cerveja. Mas é um fato interno e necessário sobre um agente que se ele julga sinceramente que x é moralmente correto, ele está motivado a perseguir o curso de ação x. Então, se um juízo moral expressa uma crença, teria de ser uma crença que sustentasse uma conexão interna e necessária com um desejo: teria de ser uma verdade necessária que um agente que possuísse uma crença teria que possuir *inter alia* [entre outras coisas] um desejo.” No entanto, nenhuma crença é necessariamente conectada a um desejo porque, como Hume sugeriu, “crenças e desejos são existências distintas” e é impossível haver uma conexão necessária entre existências distintas [...]. Então, não pode ser o caso que os juízos morais expressem crenças⁷⁰. Portanto, juízos morais não são passíveis de verdade e falsidade.

Agora nos resta saber algumas coisas como, por exemplo, que tipo de estados psicológicos os juízos morais expressam. Há alguma variedade de entendimento do que os juízos morais realmente expressam.

Ayer (1936) defendeu que os juízos morais expressam nossos sentimentos e/ou emoções. Como dissemos, se expressões de emoções/sentimentos têm a verdade variante, então elas não podem portar valor de verdade, não expressam, efetivamente, proposições morais. Na verdade, Ayer defendia que o valor de verdade de uma proposição deve poder ser verificada e, como não há como verificar o conteúdo dos juízos morais, então eles não possuem valor de verdade⁷¹. Assim, segundo Ayer, os juízos morais nada referem e, portanto, não possuem significado literal. O que queremos dizer

⁷⁰Pensamos que a melhor interpretação da expressão “os juízos morais não expressam crenças” é a de que as crenças são sempre expressas por proposições e que as proposições são as portadoras de valor de verdade, em última instância. Então, o que está sendo dito por Miller aqui é: “os juízos morais não expressam proposições, portanto não possuem valor de verdade”.

⁷¹Para mais entre emotivismo e o princípio de verificação de Ayer, ver MILLER, A. *Emotivism and the Verification Principle* In: *Proceedings of the Aristotelian Society New Series*, Vol. 98, pp. 103-12, 1998

quando expressamos um juízo moral? Os juízos morais, na verdade, seriam assim:

(a) “O assassinato é errado” = “Buuuu!!! para o assassinato/assassino” ou “=(para o assassinato/assassino”;

(b) “Ajudar os mais vulneráveis é correto” = “Aeeee!!! para os filantropos/para a ação de filantropia” ou “=) para o filantropo/para a ação de filantropia”

Ayer pensava que os juízos morais são uma espécie de externalização de um sentimento de aprovação ou desaprovação. O emotivismo de Ayer é um tanto quanto problemático, pois, para aceitarmos que juízos morais não expressam proposições genuínas, somos obrigados a aceitar toda uma teoria acerca da verdade, que é, em última instância, um tanto quanto contraintuitiva, além de termos que aceitar certa concepção de que os termos da linguagem adquirem significado literal se, e somente se, correspondente a algum objeto no mundo. Além disso, Ayer não consegue explicar como a nossa linguagem moral aparenta ter caráter realista, ou seja, nos comportamos como se ela revelasse verdades e referisse algo no mundo (ou seja, um fato moral). Em outros termos, se os juízos morais são expressões das nossas emoções, qual o sentido de discutirmos se, por exemplo, “a eutanásia é moralmente permissível?”. Do ponto de vista cognitivo (ou seja, da tentativa de revelar verdades), não há qualquer significado nesta discussão⁷².

Com um projeto mais complexo, Simon Blackburn em *Essays in Quasi-Realism* (1993) objetiva defender que os juízos morais expressam disposições (este é o elemento não-cognitivo da posição de Blackburn) em formar sentimentos de aprovação ou desaprovação perante a algo (MILLER, 2013). Além disso, e ao contrário de Ayer, Blackburn (1993) tem a preocupação central de explicar por que o nosso discurso tem a forma que tem, ou seja, esse caráter realista. Blackburn defendia uma posição chamada de projetivismo (BLACKBURN, 1993). Tal posição pode ser entendida como uma posição filosófica que defende que as propriedades de “ser bom” ou “ser correto” (o que

⁷²Para mais acerca da posição de Ayer, ver AYER, A. J. *Language, Truth, and Logic*. Nova York: Penguin Books, 2001 (first edition 1936), principalmente, capítulo VI.

Blackburn chama de “propriedades avaliativas”) são projeções de nossas emoções, ou sentimentos, ou atitudes etc. (MILLER, 2013). Dessa forma, ele pretende explicar por que nossa linguagem moral tem aqueles ares realistas que já comentamos. Reparem que Blackburn, além de querer dizer o que os juízos morais expressam, também quer explicar a aparência realista de nossa linguagem moral. Devido a essa característica é que a tese de Blackburn é chamada de quase-realismo, ou seja, analisada a fundo a linguagem moral não é realista, apesar de parecer, e expressa certas disposições humanas.

Isso posto, para quase-realistas: os juízos morais expressam coisas como:

- a “S afirma que o assassinato é errado” = “S está disposto a desaprovar qualquer ação que possa ser classificada como assassinato”, ou seja, em face de um ato de assassinato Sdesaprovaria tal ato.

O grande problema para este tipo de visão (na verdade para todas as visões não-cognitivistas) é o chamado problema Frege-Geach, que discutiremos mais adiante, pois ainda temos mais uma proposta para expor, qual seja, o expressivismo de normas de Alan Gibbard. A sua posição foi desenvolvida em seu *Wise Choices, Apt Feelings: A Theory of Normative Judgment* de 1993. Nessa obra, Gibbard defende que os juízos morais são na verdade atitudes (este é o elemento não-cognitivo) de aceitação de normas que irão governar nossas condutas e emoções. Assim,

Gibbard defende sua tese expressivista de normas. Como Gibbard salienta no prefácio de seu livro, o que o motivou inicialmente a se ocupar das questões discutidas ao longo do livro foi o problema de saber o que o termo “racional” significa. Mais do que isso, Gibbard notou que o termo “racional” (o qual ele acaba por defender que pensar que alguma coisa é racional é aceitar normas que a permitem) possui relações com todos os problemas de termos morais como “dever”, “bom” e “certo” (SANTOS & MERLUSSI, 2013 p. 346).

Assim, de acordo com Gibbard expressar que “o assassinato é errado” seria o mesmo que “dar anuência às normas que proíbem esse tipo de ato” e; podemos entender isso como: *S* é racional ao pensar que o assassinato é errado, se *S* aceita as normas de proibição de tal tipo de ato; e *S* seria irracional, se aceitasse uma norma de proibição, mas não seguisse tal norma ou não expressasse desaprovação com relação a quem a viola. Passemos ao problema Frege-Geach. Afinal, do que se trata o problema?

O ponto central do problema de Frege-Geach é que “o expressivista não pode dar uma explicação do significado das sentenças condicionais que envolvem predicados morais (bom/mau, correto/incorrecto etc) e da validade dos argumentos que envolvem sentenças morais” (D’Aversa, 2018, p. 1). Dado que a validade implica que, seja o que for que está sendo expresso nas premissas e na conclusão de um argumento, deve haver um valor de verdade possivelmente associado. Sendo assim, como podemos explicar a validade de um argumento que possua “expressões de nossos sentimentos”, como defendem os não-cognitivistas? O ponto é: se as expressões morais não possuem valor de verdade, como podemos formular [como formulamos, de fato] argumentos válidos, mas com conteúdo moral? Vamos a alguns exemplos:

Qual papel que “o aborto é permissível” desempenha no significado na seguinte condicional, dita em contexto indeterminado (expressa em P2)? (D’Aversa, 2018, p. 1)

- P1 - Os fetos não são pessoas.
- P2 -Se os fetos não são pessoas, então o aborto é moralmente permissível.
- C - Logo, o aborto é moralmente permissível.

Este argumento é válido, mas como explicar a sua validade se a consequente de P2 não expressa uma proposição? Em suma, não-cognitivistas precisam explicar como é possível que expressões morais condicionais possuem significado e, a nosso ver o mais complicado, como explicar a validade dos argumentos morais se a validade é dependente de verdade, mas as expressões

morais não possuam valor de verdade?⁷³.

Apesar de breve, nossa exposição oferece uma visão acerca das principais teses metaéticas atuais. Agora que temos uma ideia mais geral do que seja a metaética e seus problemas, assim como as suas teses, vamos passar a explorar os três grandes modelos da ética normativa, quais sejam: as éticas das virtudes, o deontologismo e o consequencialismo.

4. Ética Normativa

Assim como não é fácil distinguir a ética de outros domínios normativos, também não é simples distinguir as teorias morais normativas umas das outras. Não existe uma lista única de critérios distintivos, limitada e decisiva, que nos sirva de referência para categorizarmos todas as teorias morais normativas existentes e concebíveis. De acordo com a tradição filosófica, existem três *raciocínios morais fundamentais*: a *ética das virtudes*, centrada no aperfeiçoamento do caráter pessoal, com o objetivo de possivelmente se atingir e manter uma determinada concepção de felicidade. Depois temos o *consequencialismo*, cuja avaliação normativa se concentra na consequência das ações individuais e coletivas, sendo possível separar o consequencialismo em utilitarismo e egoísmo ético. E, por último, a *deontologia* baseada no conceito de autonomia. Além disso, existe a teoria da lei natural, que também é uma escola moral genuína, apesar de muitos a considerarem letra morta, por diferentes razões – esta não discutiremos aqui. Portanto, com o intuito de propiciar uma visão panorâmica da ética contemporânea, apresentaremos os três grandes modelos normativos fundamentais. Chamamos essas teorias de fundamentais porque elas antecedem teorias morais ou políticas derivadas como, por exemplo, a teoria do contrato social, a teoria dos direitos humanos, o liberalismo político, o utilitarismo liberal etc. que se desenvolvem a partir da conjunção das teorias do valor intrínseco com uma das teorias morais fundamentais.

⁷³Uma ótima discussão do problema Frege-Geach pode ser encontrada em D'AVERSA, R. *Moral Arguments and the Frege-Geach Problem*. PhD thesis, University of York, 2018. Disponível em <http://etheses.whiterose.ac.uk/21981/>.

4.1. Os Três Modelos Normativos

Esta introdução visa apresentar um panorama ecumênico da moralidade. Todas as escolas do pensamento moral exibem uma rica mistura de teorias da natureza humana, da noção de valor intrínseco ou bem-estar, da racionalidade, de princípios normativos e de implicações políticas. Teorias da natureza humana são, tanto quanto possível, visões meramente descritivas sobre o sujeito de nossas considerações morais, neste caso, o ser humano. Ao mesmo tempo existem concepções normativas da natureza humana. Tradicionalmente, os filósofos derivam suas teorias do valor intrínseco a partir da sua concepção de natureza humana.

Importante lembrar que, embora as escolas de pensamento moral sejam abrangentes, elas emergem de teorias ainda mais fundamentais, embora não sejam completamente redutíveis a estas teorias. Para a mitigação de possíveis dúvidas, deve-se observar que a principal função de teorias morais substanciais é a de avaliar ações, embora essa não seja uma característica que só a moralidade possua, como já apontado anteriormente. Mas as diferentes áreas de avaliação seguirão linhas muito diferentes, dependendo das várias características dos objetos de avaliação normativa, como pessoas físicas, associações de bairro, grupos financeiros, médicos, advogados, soldados, políticos etc. É essa individualização dos objetos de avaliação moral que, às vezes, determina os contornos entre a moralidade individual e pública. Porém não nos esqueçamos que a maioria de nossas ações diárias possuem consequências privadas e públicas simultaneamente. Portanto a distinção entre ética, política, e outras esferas normativas será frequentemente difícil de determinar. Assim, nosso objetivo aqui é apenas oferecer um panorama geral das escolas de pensamento moral mais influentes.

4.2. Ética das Virtudes

Todas as teorias morais oferecem uma resposta à pergunta, “o que devo fazer?”. Esta resposta é automaticamente um critério de correção ou avaliação das ações. No caso da teoria das virtudes, a resposta é “desenvolva suas virtudes”. O que torna as ações corretas é o fato de se constituírem como desenvolvimento das virtudes de um agente. A ética das virtudes, principalmente

a aristotélica, assume que os seres humanos são racionais e dependem da vida em sociedade. Adicionalmente, assume, de forma geral, uma espécie de *perfeccionismo*, a tese segundo a qual aquilo que nos beneficia não instrumentalmente é aquilo que aperfeiçoa a nossa natureza.

A princípio, a teoria das virtudes aristotélica pode parecer bastante egoísta. Mas basta nos lembrarmos de duas coisas. Primeiro, segundo a *Ética a Nicômaco*, entre o bem da cidade e o bem individual, Aristóteles diz que o primeiro é sempre supremo. Segundo, existem diversas virtudes que são sociais ou interpessoais, como ser um bom amigo, honesto, magnânimo e assim por diante. O princípio seminal da teoria das virtudes é a busca constante e ativa pela vida virtuosa. Importante frisar que a vida virtuosa não garante uma vida boa ou feliz. Para começar, normalmente, o desenvolvimento de nossas virtudes é árduo e, às vezes, até doloroso. Segundo, o exercício de certas virtudes pode exigir sacrifícios extremados e até mesmo o da própria vida do agente.

O argumento da função de Aristóteles⁷⁴ é sobre o que se deve fazer, se se quer viver bem. Ele conecta uma visão da natureza humana, ou seja, que os seres humanos são racionais, com uma visão substancial sobre o bem-estar: a vida boa é a vida da virtude. Sua concepção padrão é a seguinte:

1. O bom para algo em geral, para o que quer que tenha uma função e ação característica, depende de sua função, e o mesmo parece ser verdade para um ser humano (Livro 1, cap. 7, linha 25, 1097b)
 2. [A função dos seres humanos] é algum tipo de vida de ação da parte da alma que possui razão. A função humana é a atividade da alma de acordo com a razão. (1098a5)
 3. Virtude (o bem para alguma coisa) é exercer a própria função "bem e com perfeição" (1098a10-15)
 4. Se a função dos seres humanos é raciocinar, então ser um ser humano virtuoso é raciocinar bem. (1098a10-15)
- Conclusão: O bem humano é a atividade da alma de acordo com a virtude. (1098a15-20)

⁷⁴O argumento da função é apresentado e defendido ao longo do Livro 1 da *Ética a Nicômaco*. Aristóteles o resume no Livro 1, capítulo 7, 1097a – 1098b.

Aristóteles abre seu estudo sobre ética, mais tarde denominado *Ética a Nicômaco*, com uma investigação sobre a natureza do bem-estar humano, ou o que é uma vida boa para criaturas racionais como nós. Ele avalia o que é bom de acordo com dois testes conceituais. O bem deve ser *completo* em si mesmo, viz. desejável por si só, e não desejável por causa de algum outro bem. E o bem deve ser *autossuficiente*, viz. por si só torna a vida desejável e faltando em nada, ou seja, quando o temos, não desejamos mais nada.⁷⁵ Aristóteles chama esse estado de *eudaimonia*⁷⁶, comumente traduzido como felicidade. Estudiosos contestam se *eudaimonia* é melhor traduzida como "felicidade", "florescimento", ou simplesmente "bem-estar" ou, se deve permanecer como um termo técnico intraduzível. Escaramuças semânticas a parte, Aristóteles observa que seria enganoso pensar que a *eudaimonia* é "duradoura e que não tende a flutuar"⁷⁷, pois é óbvio que a vida tem seus altos e baixos. Então, ele pergunta: como podemos dizer que alguém está feliz, se a felicidade é uma atividade transitória da mente? E ainda, como podemos ser felizes durante a vida? Aristóteles, então, argumenta que, a fim de saber quando alguém vive bem ou o que significa que alguém vive bem, precisamos primeiro saber exatamente o que essa coisa é, seja ela uma faca, uma flauta, um soldado, um pai ou um ser humano *simpliciter*. Então, Aristóteles desenvolve uma visão metafísica segundo a qual individualizamos as coisas de acordo com sua função, ou em outras palavras,

⁷⁵1097b16.

⁷⁶Para mais informações sobre a imprecisão semântica de "eudaimonia", consulte Kraut, Richard, "Aristotle's Ethics", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/entries/aristotle-ethics/seção 2>: "Aristóteles pensa que todos concordarão que os termos "eudaimonia" ("felicidade") e "eu zên" ("viver bem") designam tal fim. O termo grego "eudaimon" é composto de duas partes: "eu" significa "bem" e "daimon" significa "divindade" ou "espírito". Portanto, significa estar vivendo de uma maneira que é favorecida por um deus. Mas Aristóteles nunca chama atenção para essa etimologia em seus escritos éticos, e ela parece ter pouca influência em seu pensamento. Ele considera o "eudaimon" um mero substituto para "eu zên" ("viver bem"). Esses termos desempenham um papel de avaliação e não são simplesmente descrições do estado de espírito de alguém.

⁷⁷Aristóteles, *Ética a Nicômaco*, (1100a-1100b, linha 35)

saber o que algo é, é saber qual é a sua função.⁷⁸ Depois que estabelecemos qual é a função de algo, passa a ser até intuitivo que este algo ter valor, ele precisa exercer bem a sua função. Quando algo exerce bem sua função, dizemos que é virtuoso.

Saber a função de algo nem sempre é tão óbvio quanto a de uma faca, uma flauta ou mesmo um soldado. E se quisermos fazer uma investigação mais generalizada e tentar dizer o que seria função única e universal de todo e qualquer ser humano, assim como a função de toda e qualquer faca? De acordo com teoria da individuação de Aristóteles, para saber a função de algo, precisamos determinar o que essa coisa faz de maneira única e nada mais pode fazer de maneira igualmente adequada. Aristóteles então conclui que a função dos seres humanos é ser racional, uma vez que a razão é o que os separa de tudo o mais, ou seja, sua singularidade. Da mesma forma, se a função dos seres humanos é ser racional, então, fazer bem é raciocinar bem. Agora, o que significa raciocinar bem para os humanos na prática? Isto jamais fica claro na obra de Aristóteles. Embora ele discursse sobre várias virtudes intrapessoais e interpessoais, não sabemos ao certo quais virtudes exercer em casos de conflito.

Sugerimos lembrar que o estado de *eudaimonia* é transitório. Então, o que nós podemos fazer é aquilo que está sob nosso controle, argumenta Aristóteles, que é permanecer em busca desse estado de vida. Neste ponto, ele conclui que a vida boa é o provável desdobramento de nossa busca constante pela virtude, a "vida ativa da virtude" em suas palavras. Portanto, em resumo, temos: 1. Ser virtuoso é exercer bem a própria função; 2. A função dos seres humanos é ser racionais. 3. Portanto, ser um ser humano virtuoso é um ser que raciocina bem. 4. Se nós raciocinarmos bem, "viveremos bem".

Note que junto com o argumento da função, há um argumento em favor do dinamismo da virtude: a boa vida requer exercício constante. Se os momentos que compõem uma boa vida são transitórios e tendem a flutuar, ficamos naturalmente interessados em saber como fazer esses bons momentos durarem mais. Na verdade, diz Aristóteles: "as mais honradas entre as próprias virtudes são mais duradouras do que outras virtudes" e segue-se que "as

⁷⁸ Aristóteles, *Política*, (1253a): "todas as coisas derivam seu caráter essencial de sua função e capacidade; e segue-se que se eles não estão mais aptos para cumprir sua função, não devemos dizer que eles ainda são as mesmas coisas."

pessoas que continuamente dedicam suas vidas a elas mais do que qualquer outra coisa” têm algum tipo de “estabilidade” (1100b15). Porque essa busca é uma atividade constante, a pessoa passa a construir e manter um certo caráter pessoal ao longo do tempo, cuja continuidade simplesmente é a vida boa. Mas como definiríamos “virtude” (*areté*, transliterado do grego)?

Aristóteles (1106b36-1107a3) defende que “a virtude/excelência é, pois, *uma escolha deliberada* relacionada a mediania, determinada pela razão e como faria alguém prudente”. Bem, já falamos sobre a importância da razão, agora é necessário falar sobre a “mediania” ou o justo meio. De acordo com Aristóteles, a virtude seria a mediania entre dois vícios – reparem que o nome pode nos enganar haja visto que não é um “meio” no sentido do meio do caminho entre dois pontos A e B. Há um componente particular relativo a nós: depende de nossa racionalidade encontrar o ponto mediano entre dois vícios. Não somos autorizados – se quisermos viver bem – a escolher um dos extremos. Vamos a um exemplo. Imagine um soldado que seja extremamente bravo e à hora da guerra corre à frente de seu pelotão. Segundo Aristóteles, ele sofre do vício da imprudência ou da temeridade, ato não recomendável. Imagine agora um soldado que, ao contrário do primeiro, deixa seu pelotão ir à frente e fica para trás, por temer a morte. Segundo a linha de raciocínio desenvolvida aqui, ele também possui um vício, qual seja, a covardia. A verdadeira virtude seria encarar a batalha com o seu pelotão, pois a união de todos os soldados em uma força tem mais sucesso na empreitada do que se houver apenas um tentando enfrentar seus inimigos ou do que se houver menos um para enfrentar tais inimigos. Então, uma pessoa verdadeiramente corajosa, seria aquela que lutaria com seus colegas ombro a ombro. Em resumo, agir seguindo o justo meio é seguir o modo correto de agir.

A última passagem da caracterização de virtude é o ponto de vista de Aristóteles de que nós aprendemos a ser virtuosos por modelos, seguindo, em última instância, o que um ser prudente faria, isto é, o que um ser que possui sabedoria prática faria – tal como denominado por Aristóteles⁷⁹. Mas o que isso

⁷⁹Segundo Aristóteles (1140b4-5) sabedoria prática é “um estado verdadeiro e racional de agir de acordo com as coisas boas ou más para o homem”. Sendo assim, pessoas com sabedoria prática sabem qual é a boa ação em uma circunstância particular, dado que conseguem perceber o justo meio.

quer dizer? Quer dizer que saber como agir de modo virtuoso é se basear no comportamento de pessoas virtuosas. Assim, a resposta para a pergunta “o que devemos fazer para desenvolver nossas virtudes?”, segundo Aristóteles seria: “faça aquilo que um virtuoso faria na circunstância em que você se encontra”. Esse critério pode parecer um tanto quanto indeterminado, afinal como saberíamos o que um virtuoso faria? Segundo Aristóteles, observando quem é virtuoso e aprendendo a partir daí. Todavia essa estratégia argumentativa parece envolver uma espécie de círculo. Se não sei de antemão quais são as atitudes virtuosas, como identificarei aqueles que as executam? Por outro lado, se sei quais atitudes são virtuosas, então por que precisaria pensar no que um virtuoso faria? Isso quer dizer que, em última instância, preciso saber quais ações são virtuosas, para saber quais ações são desta espécie

Apesar de certas dificuldades na teoria das virtudes aristotélica – como, por exemplo, seu argumento da função parecer pouco plausível para muitos, certa imprecisão sobre o que é uma virtude, ou mesmo certa falta de critério claro sobre como devemos agir –, a ética das virtudes é explorada por muitas filósofas e filósofos contemporâneos como, por exemplo, Anscombe (1958), Foot (1978) MacIntyre (1985), Zagzebski (1996), Swanton (2003), entre outras e outros que, de uma maneira ou de outra, bebem da fonte da ética aristotélica.

4.3. O Consequencialismo

Colocado de forma simples, o consequencialismo é a tese segundo a qual uma ação é correta ou incorreta dadas as suas consequências. Já as consequências são avaliadas em termos de alguma teoria do valor intrínseco. Claro que essa é a forma mais simples de enunciar a tese e que muito ainda pode ser perguntado, como as questões seguintes: (i) “qual teoria do valor intrínseco deve ser somada ao consequencialismo?” – vamos ver que, se ela variar, podemos ter teorias resultantes como o egoísmo ético e o utilitarismo; (ii) “como devemos calcular as consequências?” – considerando o montante final do valor produzido ou o valor médio dado os indivíduos?; (iii) “quais as consequências que contam: serão as atuais ou as esperadas?” – no primeiro tipo, não se considera o futuro, enquanto, no segundo, sim; (iv) “as consequências do quê?” – dos nossos atos particulares, das regras que empregamos e seguimos, das instituições sociais, dos nossos motivos?; (v) “consequências para quem?” –

para todos imparcialmente (utilitarismo), para alguns parcialmente ou para um indivíduo (egoísmo ético)⁸⁰ (SINNOTT-ARMSTRONG, 2019);

Para cada resposta diferente às questões listadas, temos uma variação de consequencialismo, dado que cada uma dessas perguntas implica uma tese logicamente independente das outras. Assim, há múltiplas teorias consequencialistas. Claramente, os dois casos paradigmáticos de teorias que aceitam o consequencialismo - e que são diametralmente opostas - são: (a) o egoísmo ético e (b) o utilitarismo clássico. Isso posto, vamos explorar com algum detalhe o egoísmo ético primeiro e depois passamos ao utilitarismo clássico.

4.3.1. Egoísmo ético

Geralmente, quando pensamos em moralidade, estamos pensando nas relações interpessoais, ou seja, no que nós devemos fazer em relação aos outros seres morais (aos outros seres humanos, ou aos animais, ou até mesmo em relação à natureza). Mas por que deveríamos pensar assim? Em outros termos, por que deveríamos ajudar outras pessoas, por exemplo? Para egoístas éticos, nosso dever moral não está relacionado com “ajudar os outros”, mas, antes, agir de tal modo a produzir as melhores consequências para nós mesmos. Portanto a tese do egoísmo ético nega que temos deveres para com terceiros, ou melhor, afirma que temos *apenas* deveres para com nós mesmos, ou ainda, que temos um único dever, qual seja, o de buscar *exclusivamente* nosso autointeresse⁸¹. Reparem na radicalidade da ideia do egoísmo ético. Tal teoria não defende que devemos buscar realizar nossos interesses enquanto realizamos os interesses dos outros indivíduos; ao contrário, devemos realizar exclusivamente nossos interesses. Os interesses dos outros indivíduos apenas são realizados se for *necessário* para realizarmos o nosso interesse. O pensamento consequencial que há no egoísmo ético é claro: o que faz de uma ação correta são suas consequências que me beneficiam, ponto final.

⁸⁰Para mais sobre o consequencialismo e suas inúmeras variações, ver SINNOTT-ARMSTRONG, Walter, "Consequentialism", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/consequentialism/>.

⁸¹A exposição aqui feita segue a estrutura da exposição de Rachels & Rachels, 2013.

Poderíamos, dessa forma, formular um princípio (regra geral através do qual todas as ações podem ser avaliadas como corretas/boas e incorretas/más), tal como o seguinte:

Princípio Egoísta: Uma ação *X* é correta se, e somente se, produz como consequência a satisfação do autointeresse de um indivíduo *I*, não sendo relevante satisfazer os interesses de terceiros, a não ser que esta ação satisfaça os interesses de *I*.

Isso posto, poderíamos genuinamente perguntar, quais as razões que teríamos para aceitar tal princípio?

Egoístas éticos defendem que a liberdade humana deve ser preservada, para que cada um possa perseguir seus planos de vida e, assim, realizar seu autointeresse. Apesar de não só o egoísmo defender que a liberdade humana é essencial para construirmos nossos planos de vida⁸², por que várias outras teorias defendem que ainda devemos ajudar os outros, em alguma medida, mas o egoísmo ético não o faça? Há vários argumentos egoístas para sustentar a tese de que não temos o dever de auxiliar os outros. Um deles é sobre o entendimento imperfeito que temos sobre a vida de terceiros, ou seja, cada um deve “cuidar da sua própria vida”, porque não sabemos como a vida de um terceiro vai. Assim, os mais bem preparados para satisfazer nosso autointeresse somos nós mesmos; e, adicionalmente, se outros tentarem nos ajudar, assim como se nós tentarmos ajudar terceiros, provavelmente, dado o entendimento imperfeito sobre a vida alheia, seríamos mais atrapalhados (ou nos atrapalhariam) do que seríamos beneficiados por tais intervenções. Então, reparem que essa é uma defesa a favor do egoísmo, tentando atacar aqueles que defendem que devemos ajudar os outros. Todavia egoístas continuam nessa linha argumentativa, defendendo que, apesar da grande possibilidade de

⁸²A teoria do valor intrínseco que o egoísmo ético aceita pode ser qualquer uma, a nível teórico. Os egoístas podem aceitar que o bem-estar é o que possui valor, eles podem aceitar que o perfeccionismo é a tese correta acerca daquilo que possui valor, entre outras. O importante notar é que o egoísmo ético possui uma forte inclinação a defender que seja o que for que seja bom em si, as pessoas devem ter a liberdade garantida para buscar tais coisas e que serem obrigadas a ajudar terceiros, pode interferir nesta liberdade.

podermos prejudicar terceiros, ao tentarmos ajudá-los, tal ajuda seria, em última instância, uma intromissão ofensiva nestas vidas. Se nós tentarmos ajudar as outras pessoas, vamos desrespeitar a autonomia dos indivíduos em buscar seu autointeresse, e isso parece ser ofensivo. Seria como se alguém dissesse a nós o que fazer, mesmo que nós não desejássemos fazer. Isso nos degradaria enquanto seres que podem fazer escolhas e que são capazes de fazê-las (Rachels & Rachels, 2013). Obviamente, essa conclusão parece inaceitável, afinal como estaríamos autorizados a nos intrometer na vida de outras pessoas, dizendo o que elas devem fazer? Isso seria de uma pretensão epistemológica enorme, ou seja, é como se disséssemos “eu sei o que é bom para você, mas você não sabe o que é bom para você mesmo”. Se supusermos que as pessoas são geralmente capazes e informadas, então a nossa pretensão de ajuda é, na verdade, uma pretensão implausível e, talvez, imoral.

Essa linha de raciocínio parece convincente, mas como apontam Rachels & Rachels (2013) o fundamento da linha de raciocínio acima não está no egoísmo, ou seja, o que está por detrás do argumento é uma preocupação outra que não a egoísta. Ao invés de uma preocupação egoísta temos, antes, uma preocupação acerca da melhor maneira de promover os melhores interesses de todos os indivíduos. Sendo assim, poderíamos nos comportar de modo egoísta, mas tal comportamento deve ser seguido não pelo princípio do egoísmo ético, mas pelo princípio da beneficência, ou seja, não interferir na vida das pessoas as beneficia mais do que o contrário. Veja como, claramente, o princípio muda. Tínhamos no começo a preocupação de apenas buscar a realização do nosso autointeresse, mas a linha argumentativa acima assume que não devemos intervir na vida alheia, porque, ao fazê-lo, estaríamos prejudicando outras pessoas. Portanto, tal linha argumentativa pode falhar para à defesa do egoísmo ético. Porém esta não é a única linha argumentativa. Ayn Rand tentou desenvolver, em seu *The Virtue of Selfishness*, de 1964, uma linha adicional que se foca na liberdade humana e no seu florescimento.

De acordo com o ponto de vista de Rand, a própria obrigação de ajudar terceiros prejudica não somente quem é ajudado, mas também que ajuda. Mas como isso seria possível? O primeiro ponto é simples. Viver sob a égide da vontade de terceiros, ou seja, da sua ajuda, é ser relegado a um estatuto de “coisa” e a uma espécie de “escravidão”. Isto degradaria o indivíduo a níveis de torná-los não livres ou, em outros termos, torna-os coisas que podemos

manipular, dada a situação em que foram colocadas. Todavia esse não é o principal problema. O problema principal é que as pessoas que são “obrigadas” a ajudar também são prejudicadas enquanto serem que desejam desenvolver e usufruir de sua individualidade. Mas, como? Ora, se fossemos obrigados a ajudar os outros, então seríamos obrigados a abrir mão de algo que não estávamos dispostos a abrir mão, como, por exemplo, nosso tempo, nossos bens materiais, nossos interesses etc. (Rachels & Rachels, 2013). Imagine a seguinte situação: um agente A decide junto a seu parceiro terem um filho e, adicionalmente, criar o filho dispendendo seus recursos para dar-lhe a melhor vida. Isso faz parte dos planos individuais dos dois. Assim, criar o filho com todos os recursos que eles possuem faz parte do autointeresse de ambos. Mas, imagine que, por alguma razão, eles são obrigados a ajudar terceiros, de tempos em tempos, ou seja, ajudar terceiros não faz parte dos interesses de ambos. Assim, quando da ajuda a terceiros, ambos terão que sacrificar algo para tanto, como, por exemplo, o tempo deles que poderia ser gasto com o seu filho ou os recursos materiais que poderiam ser gastos na criação da criança etc. Este é ponto de Rand: se somos obrigados a ajudar os outros, isso nos prejudicaria em realizar nosso autointeresse no fim das contas. E, para egoístas éticos como Rand, tal coisa é inaceitável, haja visto devermos permitir que cada *pessoa julgue a sua vida como sendo um valor último* (isso é um princípio egoísta genuíno), levando o indivíduo a sério e que isso significa não os obrigar a abrir mão daquilo que lhes interessa. Todavia essa linha argumentativa possui problemas.

O problema central que podemos elencar é o seguinte: Rand parece defender que temos duas opções de escolha, ou nós somos pessoas benevolentes e acabamos com a nossa individualidade no processo, ou nós somos egoístas éticos (Rachels & Rachels, 2013). Da forma como ela coloca as opções, soa quase irracional tentar ajudar os outros, uma vez que estaríamos escolhendo ou viver como santos morais, ou não presariamos por nossa própria vida, sujeitando-nos à “servidão moral”. Mas reparem, há uma terceira opção, qual seja, poderíamos ajudar terceiros, abrindo mão de algo supérfluo a nós (por exemplo, se temos itens demais de um mesmo produto, os itens adicionais acabam sendo supérfluos porque nós não utilizaremos de todo em todo), mas que terá impacto na vida do ajudado. E mais, o ajudado não será um mero “objeto”, se aquilo que empregarmos para ajudá-lo o faz sair de uma situação de

vulnerabilidade, tal que ele possa, posteriormente, abrir mão da ajuda. Assim, parece que Rand utiliza uma falácia do falso dilema para provar seu ponto contra quem defende que devemos ajudar os outros; e, ao fazê-lo, ela cometeria uma segunda falácia, qual seja, a do espantalho, uma vez que uma defesa mais plausível do dever de ajudar terceiros teria a forma de “devemos abrir mão do que é supérfluo” ao invés de “devemos abrir mão daquilo que é essencial para levarmos a cabo nossos interesses”. Desta forma, a posição de Rand, se as críticas forem válidas, não se sustenta. Isso posto, vamos considerar, de forma brevíssima, a posição de Thomas Hobbes a quem atribuem também uma visão egoísta.

Hobbes é conhecido por muitos como o defensor da tese de que os seres humanos *podem ser* (não quer dizer que serão)⁸³ um mal para os próprios seres humanos – quem nunca ouvi a frase “o homem é o lobo do homem”, não é verdade? Segundo Hobbes (2015), deveríamos seguir as regras morais, porque, ao segui-las, seríamos beneficiados ou, pelo menos, não seríamos prejudicados. O pensamento é o seguinte: se fizermos o que é demandado, as outras pessoas podem nos ver com bons olhos, ou nos favorecerem de algum modo, ou não nos prejudicarem, ou ficarem em dívida conosco etc. Isso iria de encontro com um interesse fundamental de não ser prejudicado (Rachels & Rachels, 2013). Essa pode ser considerada uma visão egoísta, uma vez que estaríamos olhando apenas para nós ou, em outros termos, estaríamos apenas preocupados com a nossa vida. Seguindo essa linha de raciocínio, Hobbes defendeu (2015) que o pensamento egoísta levaria à própria regra de ouro, que diz que não devemos fazer aos outros aquilo que não gostaríamos de fizessem a nós (Rachels & Rachels, 2013). Dado que se fizessemos algo aos outros, eles, provavelmente, fariam o mesmo contra nós. Bem, mas é plausível a posição de Hobbes?

Primeiramente, poderíamos colocar em causa a verdade da espécie de egoísmo psicológico que Hobbes supõe existir, mas não precisamos começar daí.

⁸³Por mais que Hobbes tenha, em alguma medida, a visão de que as pessoas seriam inclinadas ao egoísmo (ao que parece Hobbes defendia uma espécie de egoísmo psicológico), ele não descartava a possibilidade de que houvesse altruísmo, mesmo na situação hipotética do “estado de natureza”. Este “altruísmo limitado” poderia ser uma das razões para pensarmos que devemos seguir as regras morais/políticas com a finalidade de não sermos prejudicados.

A primeira observação que podemos fazer, aceitando a linha argumentativa dele é que é possível haver, pelo menos logicamente e metafisicamente, situações nas quais agir seguindo as regras morais não nos traz qualquer benefício (Rachels & Rachels, 2013). Dessa forma, se há esta possibilidade, então o argumento de Hobbes não funciona, porque, em última instância, as razões para seguirmos regras não seriam exclusivamente egoístas. O segundo ponto é que mesmo que aceitemos que somos movidos por razões egoístas em várias circunstâncias, isto não quer dizer que *sempre* ou *apenas* somos motivados por essas razões (Rachels & Rachels, 2013). Defender a tese da motivação egoísta exclusiva é, no mínimo, interpretar a psicologia motivacional humana de modo simplista e irrazoável, dado que, se alguém ajudar outra pessoa sem estar motivado por razões egoístas, a tese é refutada. Dessa forma, a posição de Hobbes não parece muito atraente.

Um problema mais geral que afeta o egoísmo ético é de que ele elege o tomador de decisão como um ser moralmente especial. O que isso quer dizer? Se defendemos que o interesse de um indivíduo sempre deve ser satisfeito independentemente dos efeitos em terceiros e o que conta para a moralidade são, portanto, as consequências para aquele indivíduo, temos o desafio de justificar porque os interesses individuais são os portadores de valor e os interesses alheios não são. Se não conseguirmos gerar uma boa justificação para tanto, estaremos sendo arbitrários e defendendo uma tese do tipo “os meus interesses é que importam, porque são meus”. Isso parece, do ponto de vista filosófico, inaceitável. O fato de certos interesses serem interesses de uma pessoa, ao invés de outra, não torna esses interesses especiais. Adicionalmente, este tipo de raciocínio é muito parecido com o raciocínio defendido por supremacistas raciais ou sexistas etc. A defesa de que alguém é moralmente especial (sejamos nós mesmos, ou um grupo específico) quebra um dos princípios mais caros à filosofia moral, qual seja, o princípio de igualdade de tratamento (Rachels & Rachels, 2013), que estabelece que não há uma razão *a priori* para tratarmos as pessoas de modo não igualitário quando estamos fazendo considerações morais. Defender que os interesses egoístas são especiais é defender que *a priori* alguém é melhor que alguém. Será, então, possível defender o egoísmo ético?

Deixamos, assim, essa reflexão aos leitores e passamos ao utilitarismo.

4.3.2. Utilitarismo

O utilitarismo clássico é, como dissemos no tópico 2, uma teoria que assume o hedonismo (seja quantitativo, seja qualitativo) como a sua visão do que é bom em si mesmo, ou seja, é uma visão acerca do bem-estar/felicidade. Porém, ele não se resume apenas a isso. Na verdade, o utilitarismo é uma teoria que assume, no mínimo, duas teses distintas, quais sejam: (i) uma tese do valor: a única coisa que possui valor em si é a felicidade/bem-estar e; (ii) uma tese do correto: as ações/regras/motivos/instituições são corretas enquanto produzem como consequência a felicidade/bem-estar. Mas, ao se tratar de uma teoria utilitarista, pelo menos as clássicas do século XIX, podemos identificar cinco teses independentes diferentes que as compõem. Quais são estas teses:

- a O Hedonismo: aquilo que porta valor em si são estado mentais. Valor positivo: o prazer; valor negativo: a dor/sofrimento;
- i O Consequencialismo: uma ação/regra é correta se, e somente se, produz as melhores consequências.
- ii O Agregacionismo: as consequências são avaliadas dado o agregado de felicidade/bem-estar que produzem. O valor pode ser “somado” para ter um total de felicidade/bem-estar descontado a possível infelicidade/mal-estar produzido.
- iii O Maximizacionismo: o importante na avaliação é que elas produzam o agregado máximo de felicidade/bem-estar, ou seja, o ato/regra deve produzir o máximo do valor, em comparação com as alternativas;
- iv A Imparcialidade/Igualdade: as ações/regras devem ser seguidas de forma totalmente imparcial, não sendo permitido tratar as pessoas de modo desigual, assim devemos ser os mais imparciais possíveis na produção do valor.

A combinação de todas estas teses nos leva a um princípio moral que pode ser formulado, de forma mais geral, da seguinte maneira:

Princípio Utilitarista: Uma ação/regra é correta se, e somente se, produz

como consequência o maior agregado possível de felicidade/bem-estar para todos os indivíduos imparcialmente.

Supondo uma formulação tão abrangente como essa, podemos identificar uma tese adicional, qual seja, a de universalidade. A universalidade significa que a teoria moral utilitarista não é relativista, ou seja, todas as teses aceitas valem para todos em todos os lugares.

Reparem na principal diferença com o egoísmo ético: a teoria utilitarista é imparcial e igualitária. Ela demanda que consideremos a felicidade/bem-estar de todos quando formos agir, e isso significa pensar de forma imparcial, ou seja, não procurar o autobenefício nem beneficiar um grupo ao invés de outro, arbitrariamente. Adicionalmente, a igualdade aqui sugerida é a de que todos os seres (até mesmo os animais não humanos) contam como iguais, já que todos são passíveis de usufruir de felicidade/bem-estar ou infelicidade/mal-estar. Vale notar que o utilitarismo consegue incorporar, tanto a forma clássica como muitas formulações contemporâneas, com considerações a animais não humanos, haja visto que tais seres são capazes de terem níveis de felicidade/bem-estar. Assim, supostamente, os animais não humanos têm que ser levados em consideração quando de uma escolha moral que irá produzir ou não felicidade/bem-estar. Portanto, na consideração das consequências dos nossos atos, deveríamos também considerar como eles impactam os animais não humanos. Em suma, a comunidade moral é estendida, para utilitaristas, não apenas a animais humanos, mas também a animais não humanos. E mais, há um requerimento de consideração igual entre estes dois grupos. Isso quer dizer que a felicidade/bem-estar de animais não humanos e de animais humanos devem ser considerados como iguais quando de uma decisão moral, não sendo os animais humanos moralmente especiais.⁸⁴

Apesar de acreditarmos que a característica (v) é compartilhada por

⁸⁴ Talvez a defesa utilitarista mais influente dos direitos dos animais seja a de Peter Singer, em *Libertação Animal (Animal Liberation)*, 2001) e *Ética Prática* (2012). Singer defende que a atitude de pensarmos que somos especiais do ponto de vista moral, em comparação com os animais não humanos, apenas por sermos humanos é uma espécie de discriminação arbitrária com os animais não humanos análogo ao ato de discriminação baseada em raça (racismo). A este tipo de discriminação chama-se de *especicismo/especismo*.

todos os utilitaristas clássicos e contemporâneos, algumas teses variam. Por exemplo, apesar de (ii) ser aceita também por todos os utilitaristas, há diferentes formulações do modelo teórico. A primeira variação é sobre *o que* deve recair a nossa avaliação, ou seja, devemos avaliar como correto ou incorreto dadas as consequências dos atos particulares de indivíduos particulares? Ou deveríamos avaliar como correto ou incorreto atos de acordo com regras que, se seguidas por quase todos, gerariam as melhores consequências? Dependendo de nossa resposta, temos ou utilitarismo de atos ou um utilitarismo de regras⁸⁵. É sobre estas variações que nos focaremos. Mas do que se trata cada uma destas formulações diferentes? Começemos por explorar o utilitarismo de atos.

Numa concepção bem simples, o utilitarismo de atos aceita uma espécie de consequencialismo chamado de *consequencialismo individual*. Segundo esta visão, para sabermos se uma ação é correta deveríamos nos perguntar: o que resultaria de uma ação particular *qualquer* se um indivíduo particular a executasse? Se resultaria o maior agregado possível de felicidade, então a ação estaria correta. Aceitando o consequencialismo individual, podemos gerar a seguinte formulação do princípio utilitarista

⁸⁵Aqui cabe um adendo sobre a nomenclatura para se referir às teses utilitaristas que serão explicadas. A nomenclatura pode variar dependendo do enfoque que damos ao critério de correção e ao processo decisório. Ao utilizarmos a nomenclatura utilitarismo de atos ou de regras, estamos olhando para o critério de correção da teoria. Dessa forma, estamos olhando o objeto a ser avaliado como correto ou incorreto, ou seja, se são atos ou se são regras. Porém, se nosso enfoque for dado ao processo decisório – como um indivíduo deve tomar suas decisões em situações morais - nós temos que mudar a nomenclatura e seguir a sugerida por Smart (1973). Smart chama o utilitarismo de atos de ‘utilitarismo direto’ uma vez que o processo decisório é o mesmo que o critério de correção, e chama o utilitarismo de atos que aceita regras de ‘utilitarismo indireto’, pois o critério de correção é diferente do processo de decisão – veremos mais sobre isso adiante. Então, do ponto de vista do critério de correção, as duas teorias podem ser reunidas sob o nome de ‘utilitarismo de atos’, enquanto se considerarmos apenas o processo decisório, teremos uma diferenciação entre o utilitarismo de atos (ou como refere Smart, utilitarismo direto) e o utilitarismo indireto. Os utilitaristas indiretos podem aceitar uma variedade de processos decisórios. Porém nós iremos discutir, mais adiante, apenas um, qual seja, o de que o melhor processo decisório é seguir regras práticas ou “aconselhativas”.

Princípio Utilitarista de Atos: uma ação *qualquer* é moralmente correta se, e somente se, executá-la gera como consequências o maior agregado de felicidade/bem-estar entre os indivíduos envolvidos.

Bem, segundo utilitaristas, o conceito de valor determina se as ações que geram um tipo de consequências são corretas. Então de acordo com o utilitarismo de atos, quando nós estivermos em uma posição de escolha entre vários cursos de ação, deveríamos escolher aquele curso de ação que mais gera o valor. Mas como fazer isso? Um modo de imaginar esse processo de decisão é à seguinte: imagine que você está perante uma situação particular qualquer X , com relação a qual você deve agir; então você se pergunta “qual curso de ação devo tomar?”. Para descobrir a resposta, você deve avaliar, primeiramente, todos os cursos de ação disponíveis naquele contexto. Depois de descobrir isso, você deve se perguntar “dado os cursos de ação disponíveis y , w e z , qual deles maximiza a felicidade/bem-estar?”. Quando determinado o quanto de felicidade/bem-estar de cada curso de ação, descontada a infelicidade/mal-estar, você pode organizar todos os cursos de ação disponíveis em uma espécie de ranking (chamado às vezes de ranking pela soma), com a ação que mais produz o valor esperado no topo, e depois a que produz menos que a primeira, e depois a que produz menos que a segunda etc. Você entendeu, não é mesmo?. Assim que o ranking for estabelecido a ação que ocupa o topo deve ser a ação a ser executada e, portanto, a ação correta. Este tipo de processo de escolha do que fazer – que chamamos, tecnicamente, de “processo decisório” – é o mais problemático neste tipo de teoria. Porém, quais são os problemas?

Os problemas relacionados ao processo decisório desta forma mais simples de utilitarismo de atos são inúmeros. O primeiro deles é que o processo é epistemicamente custoso, em outros termos, para tomarmos uma decisão, devemos conhecer muitas coisas, como, por exemplo, todos os cursos de ação disponíveis, como cada curso de ação afetará cada indivíduo envolvido, a extensão do conjunto de indivíduos envolvidos (quantos são afinal? A quantidade é relevante para a soma do agregado), etc. Dada a complexidade dos ambientes de escolha, por vezes seria impossível fazer esse tipo de levantamento, principalmente, quando o tempo é um fator relevante para tomada de decisão. Em algumas situações, não teríamos tempo de calcular e sobrepesar todos os cursos de ações disponíveis. Nelas as ações devem ser tomadas de modo

rápido, de modo que seria impossível seguir o princípio utilitarista de atos, já que não seria possível aplicá-lo nessas situações. Por vezes, escolher uma ação rapidamente é crucial para que possamos atingir certos objetivos, e não outros; mas fazê-la com o rigor exigido pelo princípio se torna muito difícil. Podemos agir de forma rápida, mas apenas gerar resultados sub-ótimos, ou seja, aqueles que geram boas consequências, mas não o máximo possível delas. Quando o próprio tempo é peça chave na geração ou não das melhores consequências, não poderíamos seguir todo aquele processo de decisão defendido pelo utilitarismo de atos, no final das contas.

O segundo problema mais gritante com esta forma de utilitarismo de atos são os chamados problemas de cooperação. Em alguns contextos, um indivíduo sozinho não conseguiria gerar as melhores consequências; ele necessitaria da cooperação dos outros indivíduos, para atingir o objetivo. Imagine, por exemplo, que não há qualquer regra sobre qual lado da estrada dirigir. Você pode dirigir pelo lado direito ou pelo lado esquerdo, seguindo o que você pensa ser o melhor. Todavia, para que o trânsito flua bem e não cause acidente, ou seja, para que o trânsito cause mais felicidade/bem-estar que o contrário, seria necessário a coordenação das pessoas, para que elas façam o mesmo. Mas, se seguirmos o princípio utilitarista de atos e, portanto, fosse deixada a cargo de cada indivíduo a escolha sobre qual lado dirigir, poderíamos ter uma ampla discordância entre todos e, como consequência, nosso trânsito seria muito perigoso, com cada um dirigindo pelo lado que julga melhor. Assim, não seria aconselhável seguir o princípio estabelecido acima. O que esses problemas apontam é que calculadoras constantes das melhores consequências não conseguem (ou falham na maioria das vezes em) gerá-las. Se o critério de correção de uma ação moral também é o processo decisório individual, então as pessoas seriam obrigadas a calcular os resultados de cada ato, de modo que, para grande parte deles, elas não conseguiriam cumprir com o critério de correção. Sendo assim, alguma coisa no processo decisório precisaria ser reformulada. Já veremos como isso pode ser feito, mas antes, vamos mencionar um segundo conjunto de problemas, quais sejam, os problemas relacionados a justiça.

Geralmente, pensamos que seja qual for o princípio moral que estamos admitindo ele deve conservar a justiça ou, pelo menos, aquilo que é intuitivamente justo. Parece justo que, por exemplo, respeitemos os direitos de

terceiros. Assim, pareceria injusto que nós agíssemos de tal forma a desrespeitar os direitos individuais. Imagine que, para que você consiga maximizar o bem-estar, você precise desrespeitar um direito, digamos o direito à integridade física de um único indivíduo. Será que deveríamos realizar tal ato? Se a sua intuição aponta para a resposta “não, não devemos desrespeitar o direito à integridade física de ninguém”, então você acabou por descobrir que “gerar o máximo de felicidade/bem-estar” e “respeitar direitos” não se conciliariam em alguns contextos. Em outros termos, temos requerimentos morais de naturezas diferentes, que não podem ser satisfeitos simultaneamente e, se escolhermos ficar com o nosso princípio utilitarista de atos, tornamo-nos injustos; mas, se escolhermos respeitar direitos ao invés de seguir o princípio – supondo que já o aceitávamos antes – então somos inconsistentes, dado que, em cada contexto, utilizaríamos um princípio que elege a felicidade/bem-estar como algo bom ou um princípio que elege o respeito aos direitos como algo bom. Resumindo, a exigência do princípio utilitarista de atos, de geração das melhores consequências ato a ato, é diferente da exigência de respeitar direitos e que, em várias circunstâncias, seguir um, viola o outro. Há como resolver os problemas citados, para que continuemos como um utilitarista de atos?

Para continuarmos com o utilitarismo de atos e não sofrermos com tais problemas, podemos defender que o processo de tomada de decisão deve ser diferente do princípio utilitarista de atos. O que isso quer dizer é que por mais que continuemos a defendê-lo, segui-lo estritamente não gera as melhores consequências – como vimos no problema do tempo, da cooperação e da justiça. Assim, pode-se propor outro processo decisório, qual seja, que devemos seguir as regras da moralidade comum na maior parte das vezes como regras “aconselhativas” para a tomada de decisão. Assim, não precisamos nos perguntar se um curso de ação que envolve o assassinato gerará ou não a maior soma de felicidade/bem-estar, devemos apenas seguir a regra comum de proibição do assassinato e pronto. Essa forma de utilitarismo de atos é, por vezes, chamada de utilitarismo de atos do nível múltiplo (CRISP, 2006) e é atribuída a John Stuart Mill (2000a). De forma mais clara, esta forma estabelece um princípio utilitarista um pouco mais complexo e pode ser formulado como se segue:

Princípio Utilitarista de Atos de múltiplos níveis: = (a) uma ação *qualquer*

é moralmente correta se, e somente se, executá-la gera as melhores consequências para todos os indivíduos naquele contexto e, (b) geralmente, a melhor forma de gerá-las é seguir as regras da moralidade comum.

Reparem no “geralmente” no ponto (b). Ele estabelece que de forma geral o processo decisório é melhor se seguirmos as regras da moralidade comum e não que *devemos* fazer isto sempre. Isso é relevante, pois, algumas regras da moralidade comum não maximizam a felicidade/bem-estar em alguns contextos. Sendo assim, nos contextos que percebemos que seguir as regras da moralidade comum não irá maximizar a felicidade/bem-estar, podemos abandoná-las. Bem, esta estratégia parece resolver alguns dos problemas acima. Seguir as regras da moralidade comum, nos poupa tempo e esforço epistêmico, também nos permite cooperar. Mas por que deveríamos aceitar uma mudança no processo de decisão? Não seria arbitrário aceitar a modificação apenas porque ela dá menos problemas? A modificação é baseada em razões que o próprio utilitarismo pode apresentar. Por exemplo, se somos utilitaristas e estamos comprometidos a gerar o máximo de felicidade/bem-estar e, visto que o processo decisório do utilitarismo de atos simples não gera o máximo de felicidade, então nada mais razoável para quem está comprometido do que a mudança do processo de decisão. Isso quer dizer que, dado o fim relevante, parece que estamos justificados a modificar o nosso processo de decisão. Mas e a justificação do respeito aos direitos?

O utilitarismo de atos que aceita regras não parece solucionar bem o problema dos direitos. Alguém poderia argumentar que é uma regra da moralidade comum “respeitar direitos”, ou seja, parece fazer parte da nossa ideia de moralidade o respeito aos direitos e, portanto, não sairíamos por aí desrespeitando-os. O problema é que a moralidade comum estipula regras que nem sempre estão de acordo com o que parece correto. Alguns direitos atribuídos aos indivíduos podem não ser genuinamente direitos. Por exemplo, alguém poderia ter o “direito” ao poder absoluto sobre a população, ou uma suposta casta superior ter o “direito” de dominar casta consideradas inferiores; mesmo que isso faça parte da moralidade comum de um povo, não parece justo. Não poderíamos respeitar esses direitos e seguir os “deveres” que eles estabelecem. Tais “direitos” deveriam ser abandonados de início e nunca ser

aplicados, uma vez que a existência de deles parece gerar piores consequências do que as melhores ao longo do tempo. Assim, não há apenas o problema de respeitar direitos, mas também o de saber que direitos temos. Em democracias, parece mais fácil dizer que faz parte da moralidade comum respeitar direitos, mas fora delas a própria atribuição de direitos segue critérios arbitrários que vão contra as nossas intuições mais básicas acerca da igualdade, por exemplo. Então, se o utilitarismo de atos que aceita considerações prudenciais sobre as regras da moralidade comum não consegue resolver o problema dos direitos, deveríamos abandonar o utilitarismo?

Para alguns pensadores (BRANDT, 1979, HOOKER, 2000, por exemplo), nós não deveríamos abandonar o utilitarismo; nós deveríamos apenas modificá-lo, sem que ele perca a essência. Mas qual a essência da teoria? A essência é a preocupação de que a geração de felicidade/bem-estar através das consequências é algo primordial em moralidade. Sendo assim, a modificação que vamos explorar agora é o que se chama comumente de utilitarismo de regras⁸⁶. O utilitarismo de regras assume várias formulações. Mas aqui vamos tentar explorar duas formulações: (a) o utilitarismo de regras do código moral vigente e (b) o utilitarismo do código moral ideal. Mas, antes vamos fazer algumas considerações gerais sobre a estrutura do utilitarismo de regras.

Ao contrário do utilitarismo de atos (seja qual for o tipo), o de regras encara a moralidade como um empreendimento coletivo. Em termos mais técnicos, o utilitarismo de regras (seja qual for o tipo) aceita o que chamamos de *consequencialismo coletivo*. Segundo esta visão da moralidade, para sabermos o que deveríamos fazer, deveríamos nos perguntar “e se todos fizessem X, o que resultaria disso?” A resposta a essa pergunta nos levaria a pensar a moralidade como um empreendimento coletivo e selecionar regras que se seguidas por todos os competentes, produziria as melhores consequências. Para frisar mais as diferenças entre os dois grandes tipos de utilitarismo (o de atos e o de regras), podemos observar que para o utilitarismo de atos que aceita regras, elas

⁸⁶Há vários tipos de consequencialismo que são aceitos por teorias utilitaristas de regras, como, por exemplo, o consequencialismo de regras de generalização, consequencialismo de regras institucional (GOODWIN, 1995), consequencialismo de regras do código moral vigente (BRANDT, 1992), consequencialismo de regras do código moral Ideal (HOOKER, 2000, SANTOS, 2017) etc.

somente desempenham papel no processo de decisão, ou seja, elas são aconselhativas. No utilitarismo de atos simples, elas não desempenham papel algum. Já no utilitarismo de regras, elas desempenham papel central, não apenas no processo de decisão, mas também no critério de correção. O que isso quer dizer? Que para o utilitarismo de regras uma ação é moralmente correta se, e somente se, ela respeita uma regra que, se todos seguissem, geraria a maximização de felicidade/bem-estar. Assim, podemos expressar um princípio do utilitarismo de regras como se segue:

Princípio Utilitarista de Regras: uma ação é moralmente correta se, e somente se, respeita uma regra que geraria, quando seguida por todos os indivíduos, o maior agregado de felicidade/bem-estar como consequências, de modo imparcial.

Assim, o utilitarismo de regras modifica apenas o consequencialismo aceito, *podendo* – mas não necessariamente – manter as outras características como o maximizacionismo, o agregacionismo, a imparcialidade etc. Essa ideia surge não apenas para resolver os problemas dos tipos de utilitarismo de atos, mas também por causa de um entendimento de que a moralidade deve ser um empreendimento cooperativo por essência. Todavia, podemos nos perguntar: quais regras deveríamos seguir?

No utilitarismo de regras do código moral, a defesa é que deveríamos aceitar um código de regras que, se seguido por todos os indivíduos aptos, maximizará a felicidade/bem-estar. Então, o primeiro desafio é entender como selecionamos esse código. Deveria ele ser o que já está operando em uma determinada sociedade? Ou ele deve incluir regras que não são integrantes de qualquer cultura, por exemplo? Se respondemos afirmativamente para a primeira pergunta temos o utilitarismo de regras do código moral vigente (doravante, URCMV). Por outro lado, se pensamos que devemos construir um código que também incluía regras não existentes atualmente, então iremos defender um tipo de utilitarismo de regras do código moral ideal (doravante, URCMI). A grande diferença entre ambos, portanto, é qual o código que deveríamos seguir. Os defensores do URCMV defendem que o código correto deve ser aquele que já faz parte da nossa cultura, enquanto os defensores do URCMI defendem que o melhor é o código que também possui regras não existentes atualmente. Para

deixar mais claro, vamos explorar cada uma das propostas, mas como sempre, esta é uma brevíssima apresentação.

O URCMV defende que a resposta para a pergunta “se todos fizessem X, o que resultaria?” seria que a *corretude de X* dependerá do código dado em uma sociedade. Richard Brandt estabelece a posição como se segue:

Consequencialismo de regras pode ser dividido em dois grupos principais, de acordo com a corretude de um ato particular é feita uma função de regras ideais em algum sentido, ou regras vigentes e reconhecidas em uma sociedade. A teoria variante que eu explicarei é a do primeiro tipo. [...] De acordo com o segundo tipo de teoria, um dever ou obrigação de um indivíduo em uma situação particular é determinado, com algumas exceções, somente por regras morais, ou instituições, ou práticas prevalentes na sociedade, e não por quais regras (etc.) seria idealmente melhor possuir na sociedade. (Às vezes, é defendido que as regras morais vigentes, práticas etc. são apenas uma condição necessária para um ato ser moralmente obrigatório ou errado) (BRANDT, 1992, p.115).

Assim, sendo um código moral vigente é aquele composto por regras que já estão arraigadas em uma forma de vida social e, portanto, a corretude de uma ação é definida enquanto ela se coaduna com as regras do código moral. Isso posto, podemos formular um princípio para URCMV como segue:

Princípio do URCMV: Uma ação X é moralmente correta se, e somente se, for permitida por um código cujas regras são comumente aceitas pelos indivíduos dentro de uma sociedade e produzem as melhores consequências.

Duas coisas estão implícitas neste tipo de pensamento, além é claro da ideia de que a moralidade é um empreendimento coletivo, quais sejam: que a moralidade comum preserva regras que maximizam a felicidade/bem-estar, se

forem seguidas por todos, e que haverá amplo endosso do código dentro da sociedade. A segunda ideia implícita está relacionada com o fato de que a sociedade já aceita o código e isso implica que os custos de ensinar às pessoas o código e os custos para a sua internalização são bem baixos em termos de felicidade/bem-estar. Mas o que isso quer dizer? Quando aceitamos certas regras e as internalizamos, isso quer dizer que formamos crenças e hábitos de ação baseado nelas. Por vezes, tais crenças e hábitos estão tão arraigados em como vivemos nossas vidas, que tentar mudá-los causaria custos em termos de bem-estar, ou seja, nós teríamos um profundo mal-estar, porque teríamos de rever nossas crenças e hábitos, e isso envolve mudar nossa visão de mundo, o que, por sua vez, envolve abrir mão de coisas que pensávamos estarem corretas e/ou que beneficiavam a nós e aos outros não instrumentalmente. Se for verdade, então o URMCV tem uma vantagem muito boa, o agregado de felicidade/bem-estar será maior dado que não causaremos infelicidade/mal-estar nas pessoas – lembrem-se que o agregado sempre tem de levar em conta a parte negativa (a infelicidade/mal-estar), para ser definido. Mas alguém poderia argumentar que mesmo que o código seja vigente, nós teremos aquelas pessoas que não o seguem. Bem, isso é verdade, mas aqui entra o segundo ponto positivo da teoria. O pressuposto de que haverá amplo endosso não é arbitrário, pois, se a maior parte das pessoas endossa o código, então, muito provavelmente, elas irão funcionar como contrapeso àqueles que não endossam, forçando-os a seguir as regras. Isso quer dizer que a teoria oferece força motivacional externa aos indivíduos. Assim, mesmo que alguém na sociedade não endosse que temos que dirigir pelo lado direito, nós o forçamos a isso, e não apenas porque, se ele não o fizer, nós iremos puni-lo (com multas, por exemplo), mas porque nós iremos organizar o funcionamento do trânsito desta maneira e ele não terá escolha a não ser seguir dirigindo pelo lado direito – claro, supondo que ele é um indivíduo racional e não quer se colocar em risco de morte.

Um dos grandes problemas desse modelo teórico, porém, está na ideia implícita de que as sociedades perpetuarão apenas regras que maximizam a felicidade/bem-estar, o que não é verdade. A depender da sociedade, algumas regras morais podem ser aceitas, estar profundamente arraigadas na forma de vida social e, mesmo assim, parecer indesejáveis do ponto de vista da geração do bem-estar. Imagine, por exemplo, uma sociedade que é dividida em castas, na qual as castas superiores possuem uma série de regalias e direitos especiais,

enquanto as castas inferiores estão fadadas a viver na miséria e com um conjunto de direitos mínimos (ou mesmo sem direitos). Será que um código que aceita regras que dão manutenção a este tipo de coisa é o melhor código a ser seguido do ponto de vista da promoção do maior agregado possível de felicidade/bem-estar? Não há qualquer outro código alternativo que faça um trabalho melhor? A resposta parece ser, “sim, há outros códigos que fazem um trabalho melhor em termos de maximização do bem-estar”. O problema central é que o código parece não revelar aquilo que pensamos ser intuitivamente correto neste caso. Os códigos vigentes estão impregnados com crenças falsas, valores distorcidos e preconceitos, não importando sobre qual sociedade estamos falando⁸⁷. Isso parece tão problemático, que leva a nos movermos para o URCMI.

O URCMI defende, grosso modo, a seguinte tese: o melhor código a ser seguido e aceito por todos os indivíduos competentes, ou seja, aqueles que tem a capacidade de avaliar e anuir com o código, é aquele que promove a maximização da felicidade/bem-estar e é composto por regras que não necessariamente fazem parte do modo de vida de uma sociedade. Reparem que o “ideal” vem justamente desta característica, qual seja, as regras que compõem o código podem ser inexistentes no mundo atual. Sendo assim podemos formular um princípio mais geral do URCMI:

Princípio do URCMI: um ato é moralmente correto se, e somente se, ele é permitido/exigido por um código moral que, se fosse aceito pela maioria das pessoas e se pudesse ser ensinado para cada nova geração, produziria as melhores consequências.

Assim, ao selecionarmos o código que deve ser aceito e seguido, temos de analisar os códigos disponíveis e selecionar aquele que resultará na

⁸⁷Antes que se pense que a teoria do URCMV não é adequada, por não ser universal no sentido de que o que ela estabelece como correto não é válido para todas as sociedades; uma vez que os códigos irão variar de sociedade para sociedade, uma observação pode ser feita: a teoria não tem essa pretensão. Por exemplo, a teoria da justiça de Rawls apresentada em *Uma Teoria da Justiça*, por exemplo, não pode ser acusada de não ser universal, uma vez que ela é construída tendo como objeto apenas as sociedades democráticas. O fato de a teoria não ser universal ou ser restrita não é um problema em si.

maximização da felicidade/bem-estar, mesmo que nele ocorra várias regras ideais. Os passos que devemos dar são, portanto, (i) selecionar o código dentre todos os alternativos que gerará o máximo de felicidade/bem-estar e; (ii) avaliar as ações como corretas ou incorretas em termos de conformação delas as regras. O principal motivo para aceitarmos essa estratégia é a ideia de que os códigos vigentes são falhos em termos de produzir as consequências que esperamos que deles se produza. Além disso, o URCMI implica algo como o melhoramento moral, ou seja, somos capazes de avaliar as regras existentes, identificar aquelas que parecem indesejáveis e substituir por aquelas que nos parecem melhores. Portanto, em comparação com o URCMV, nós temos essas características como positivas: podemos ter regras melhores em termos de maximização do bem-estar e alguma ideia de progresso moral.

Como em toda teoria filosófica, também há perdas, se aceitamos o URCMI. Por exemplo, os problemas dos custos de internalização são gigantescos, pois se exigirmos que as pessoas que vivem agora internalizem o código, elas terão que mudar suas vidas, nos termos que explicamos acima, e isso diminui nosso montante final de bem-estar. Outro problema, este mais abstrato, é o de saber quais os possíveis códigos disponíveis. Como os códigos podem assumir regras ideais, deve haver um grande número de códigos a serem considerados e, portanto, a nível prático, parece não ser possível considerar cada um deles. Há vários outros problemas⁸⁸ que afetam esta versão do

⁸⁸São três os conjuntos de objeções que já são consideradas clássicas contra o URCMI, quais sejam: (1) *A objeção do Colapso*: o consequencialismo de regras do código moral ideal colapsa no consequencialismo de ato; então o consequencialismo de regras do código moral ideal é descartável, pois não nos apresenta uma teoria distinta do consequencialismo de ato do nível simples ou do consequencialismo de ato de nível múltiplo; (2) *A objeção da adoração à regra*: o consequencialismo de regras defende que devemos avaliar as regras por sua capacidade de produzir as melhores consequências e requer que a sigamos, mas, ao nos depararmos com uma situação particular na qual seguir a regra não produz as melhores consequências, ele demanda que abandonemos a regra (o que faz a teoria incoerente) ou nos diz para seguir a regra (o que torna a teoria implausível, dado os maus resultados que seriam gerados). (3) *Objeção da Aceitação Parcial*: se o consequencialismo de regras do código moral ideal estabelece que nós possuímos responsabilidades compartilhadas e que a maximização apenas ocorre

utilitarismo, mas teremos que deixá-los de lado, para passarmos à próxima teoria normativa considerada.

Agora que conhecemos um pouco sobre o consequencialismo, o egoísmo ético e as variações do utilitarismo, podemos passar a apresentar e considerar uma das teorias morais mais influentes, qual seja, o deontologismo kantiano. Immanuel Kant é tão influente, que até hoje suas ideias principais são amplamente defendidas. De um modo curioso, o deontologismo kantiano deve ser o modelo teórico mais popular na academia brasileira. Sendo assim, vamos passar a apresentar o modelo desenvolvido por ele.

4.4. Deontologia Kantiana

Quando consideramos o modelo deontologista na ética, temos que ter em mente de qual de suas variações estamos lidando, todavia, sem dúvidas, o principal modelo deontologista é de Immanuel Kant (alegadamente, o fundador do deontologismo). O termo “deontologia” vem da união de dois termos gregos “*deontos*” que significa “dever” e “*logos*” que significa algo como “razão”; assim, costuma-se entender que o termo “deontologia” refere-se ao estudo dos deveres. E foi justamente isso que Kant fez em seu livro *Fundamentação da Metafísica dos Costumes* (2009). Assim como éticas utilitaristas, o deontologismo kantiano também pretende estabelecer um princípio moral que nos ofereça uma visão clara daquilo que devemos fazer. Kant defendia certos aspectos que são pressupostos na sua defesa da moralidade, qual sejam: seres humanos são racionais e, portanto, livres para fazer escolhas. Tal racionalidade nos faz autônomos, ou seja, pessoas que podem fazer escolhas por si mesmo (sem a interferência de terceiros) e decidir as próprias regras de conduta, distinguir informações relevantes, ser capazes de obter conhecimento etc. Segundo Kant, a autonomia e a racionalidade nos conferem “dignidade”, uma característica moral relevante. Então, segundo o filósofo alemão, nossa autonomia possui valor

quando todos os indivíduos do grupo aceitam e cumprem tais obrigações, quais seriam as nossas responsabilidades se grande parte dos indivíduos não aceitam o código de regras ideal? Para mais sobre a objeção do colapso ver LYONS, D. *Forms and Limits of Utilitarianism*. Oxford: Oxford University Press, 1965, p.119s.

intrínseco e deve ser respeitada – analogamente a utilitaristas que defendem que a felicidade/bem-estar possui valor intrínseco. O valor fundamental da ética kantiana seria a nossa dignidade. Ela pode ser entendida como dando a nós valor intrínseco, em outros termos, cada pessoa possuidora de racionalidade e autonomia tem valor em si e não pode ser desrespeitada, nem mesmo para a produção da felicidade/bem-estar de todos os indivíduos. Seres humanos são passíveis de consideração moral, dado possuírem autonomia e racionalidade. Pelo menos do ponto de vista kantiano, a comunidade moral apenas é formada por seres que são autônomos e racionais, portanto a deontologia kantiana deixa de fora da comunidade moral todos os animais não humanos. A ênfase kantiana nesta propriedade moral relevante, como veremos ao final desta seção, trará o problema de que a tese kantiana implica que animais não humanos não são passíveis de consideração moral direta, mas deixemos isto para o final de nossa apresentação.

Kant tentou estabelecer seu princípio moral fundamental a partir da análise do discurso moral cotidiano, fazendo notar que há duas maneiras, pelo menos, em que utilizamos o termo “deve/dever”. Assim, temos que diferenciar entre dois “sentidos” de dever: (a) Um uso baseado em nossos desejos e; (b) Um uso nunca baseado em nossos desejos. Sobre (a), Kant defendeu que temos certos “deveres” dado nossos desejos específicos, a tal teste de deveres deu o nome de Imperativo Hipotético (I.H.). O nome parece esquisito, mas faz todo o sentido, por exemplo, dado a *hipótese* de que você deseja X, então você *deve* procurar pelos meios relevantes para alcançar X (Imperativo/ordem). Assim, dado um desejo específico, temos uma ordem (ordem no sentido de mando) específica. O teste do Imperativo Hipotético estabelece o que “devemos” fazer dado um desejo prévio, mas não estabelece qualquer dever (obrigação moral) no sentido forte, uma vez que o que devemos fazer dado o I.H. depende de nossos desejos. Troque o desejo, trocará o que se deve fazer. O I.H. segue o que, comumente, se chama de racionalidade pragmática: se objetiva fazer X, então use os melhores meios para fazê-lo.

Já sobre (b), Kant defendeu que nossas obrigações morais não dependem *nunca* do que desejamos. A forma de pensamento não é mais: “se quiser X, então deve fazer Y”. Os deveres morais são categóricos, e o pensamento expresso é: *você deve fazer assim e de nenhuma outra maneira, independente do que você deseja*. Ao analisar este uso do conceito de “dever”,

Kant fez notar que deveria haver um princípio fundamental da moralidade que, através do uso da nossa racionalidade, nos permitiria definir todos os nossos deveres. Ele chamava de Imperativo Categórico (I.C). O filósofo pensava que o I.C revelava sempre o que era racional de se fazer, portanto nossos deveres estariam baseados em exigências da racionalidade e não em qualquer entidade divina. Ele argumentava que poderíamos decidir quais eram nossos deveres, usando o raciocínio moral, através de um teste, usando o I.C. Como dito acima, Kant defendia que os seres humanos são os únicos possuidores de valor intrínseco, isso quer dizer que nós nunca deveríamos tratar as pessoas como “objetos”, ou seja, nós nunca deveríamos utilizar as pessoas como meros meios para realizarmos aquilo que desejamos⁸⁹, dado que isso seria um profundo desrespeito àquilo que é valioso por si mesmo e, segundo Kant, dado que nós somos tão valiosos, a moralidade requereria que sempre fôssemos respeitados. Assim sendo, Kant cunhou seu princípio moral fundamental, o Imperativo Categórico (há várias formulações diferentes do princípio na obra kantiana, portanto utilizaremos apenas uma)⁹⁰ que em sua formulação geral, pode ser

⁸⁹É importante fazer uma distinção sobre como devemos entender “utilizar uma pessoa como um mero meio” e “utilizar as pessoas como meio de modo respeitoso”. Usar uma pessoa como um mero meio é tomá-la como um objeto, ou seja, não se importar com as escolhas autônomas delas; assim, se enganamos uma pessoa para ganho próprio estamos utilizando-a como mero meio; se restringimos a sua liberdade de ação de modo arbitrário, sem o consentimento dela, também estamos utilizando-a como mero meio etc. e, segundo Kant, isso é imoral. Por outro lado, é comum utilizar as pessoas como meio. Por exemplo, um aluno utiliza um professor para obter conhecimento, um paciente utiliza uma médica para obter o reestabelecimento de sua saúde etc., porém isto não é imoral dado que tanto o professor como a médica decidiram autonomamente e de forma racional entrarem nestas profissões e sabiam de antemão que teriam alunos e pacientes, por exemplo. Assim, usar como mero meio é não considerar as escolhas autônomas ou prejudicar tais escolhas (com ludibriação, por exemplo); usar como meio é respeitar as escolhas autônomas dos indivíduos que racionalmente e livremente decidiram por fazer o que fazem.

⁹⁰As outras formulações que podem ser encontradas na *Fundamentação da Metafísica dos Costumes* (Grundlegung zur Metaphysik der Sitte (GMS), em alemão) são as que se seguem: (a) “Age como se a máxima de tua ação se devesse tornar, pela tua vontade, em uma lei universal da natureza” (GMS 4:421) – chamada pelos estudiosos da obra de Kant de “formulação da lei da natureza”; (b) “Age de tal maneira que uses a humanidade,

expressa da seguinte forma:

Imperativo Categórico: “Age apenas segundo a máxima tal que possas ao mesmo tempo querer que ela se torne uma lei universal” (GMS, 4:421)

Vamos tentar entender cada elemento teórico que está presente nesta formulação tais como “máxima” e “lei universal”.

Uma “máxima” é uma regra subjetiva de ação, ou seja, é uma regra de ação que um indivíduo autônomo particular segue em sua vida particular. Então, por exemplo, se um indivíduo *I* segue a regra “não devo mentir” em sua vida, esta é uma máxima. Simples assim. Já uma “lei universal” é uma regra que é válida a todos os indivíduos em todos os lugares e em todos os tempos. Sendo assim, o “teste” do imperativo categórico nada mais é do que uma regra que nos permite avaliar se regras subjetivas podem ser regras universais e, portanto, definir deveres. O teste do I.C. estipula se é racional aceitarmos que uma máxima *X*, possa ser aceita racionalmente como guia de ações, universalmente. Com todos os seres humanos racionais podem utilizar o teste e constatar se uma máxima pode ser aceita como lei universal, teríamos como resultado um dever que pode ser aceito de forma racional por todos.

Façamos um teste: tomemos a regra da proibição da mentira. Será que ela pode ser aceita universalmente, dada a consideração de que as pessoas são seres valiosos por si mesmos, pela sua dignidade? Pensem da seguinte forma: “e se fosse permitido às pessoas mentir, o que resultaria em termos de respeito à autonomias?”. Ao que parece, quando mentimos para alguém, estamos interferindo em suas escolhas autônomas, dado que a escolha do indivíduo será baseada agora em uma informação falsa, que pode ter como objetivo levar esta pessoa a pensar e escolher o que o mentiroso deseja. Isso parece um claro

tanto na tua pessoa como na pessoa de qualquer outro, sempre e simultaneamente como fim e nunca simplesmente como meio” (GMS 4:429) – chamada de “formulação da humanidade; (c) “Age de tal maneira que a tua vontade pela sua máxima se possa considerar a si mesmo ao mesmo tempo legisladora universal” (GMS 4:434) chamada de “formulação da autonomia”. A certa discussão se todas as formulações são coextensivas ou não, ou seja, se todas elas dizem o mesmo ou não.

desrespeito a autonomia individual, por usar uma pessoa como se fosse uma peça em um grande jogo formado pelo mentiroso, um meio para alcançar suas finalidades. Assim sendo, a regra de permissão da mentira não pode ser aceita; por contraposição a negativa dela pode ser aceita, não mentir é respeitar os indivíduos como fins em si mesmo, portanto, a regra “não devemos mentir” pode ser considerada uma lei universal válida para todos e, portanto, um dever. Ótimo, conseguimos estipular deveres que qualquer ser humano racional que reconhece as pessoas como fins em si mesmas aceitaria. Sendo assim, se uma pessoa não mente, ela está agindo moralmente, certo? Não, não é bem assim do ponto de vista kantiano, ainda nos falta um elemento, a motivação correta.

Kant defendia que uma ação não é correta meramente porque alguém segue a lei universal, mas porque as pessoas a seguem *por reconhecer que aquilo é que deve ser feito*. Kant chamou a motivação de *agir por dever*, ou seja, por aceitar racionalmente que algo deve ser feito e que não poderia ser diferente, de *vontade boa*. Quando alguém segue uma regra moral (uma lei universal), por querer ser visto como uma pessoa boa ou parar tirar alguma vantagem, ela está agindo imoralmente. Por isso que se diz que a ética kantiana tem a motivação como um dos elementos centrais para a definição de se uma ação foi moralmente correta ou incorreta. Para deixar isso mais claro, considere o utilitarismo em comparação ao deontologismo kantiano. Para utilitaristas clássicos (algumas formas de utilitarismo contemporâneos pensam que a motivação tem relevância), pouco importa se uma pessoa está gerando as melhores consequências para se autopromover, por exemplo; o que define que a ação foi correta são única e exclusivamente as consequências. Já para o deontologismo kantiano é o oposto, uma ação é moralmente correta se o indivíduo a executa motivado pelo reconhecimento de que aquilo que ele faz deve ser feito porque é seu dever, não importando as consequências. Assim, para o deontologismo kantiano o consequencialismo é falso, pois pouco importam as consequências para definição de se uma ação é moralmente correta; o que importa é a motivação correta. E, ainda mais forte, uma lei universal, ou seja, uma regra moral válida para todos, não pode ser violada ou desrespeitada, não importando as consequências geradas ao segui-la. Por esta razão é comum ouvirmos que o deontologismo kantiano é avesso ao consequencialismo.

Se pudéssemos resumir os elementos principais da teoria kantiana,

poderíamos elencar em pelo menos três pontos: (i) o valor moral de uma ação está em seu cumprimento por dever (e não meramente em conformidade com o dever), ou seja, uma ação possui valor moral apenas se o tomador de decisão estiver agindo motivado pela razão correta; (ii) uma ação desempenhada por dever **não** tem seu valor dadas as consequências que gera ou de acordo com as virtudes do indivíduo que toma a decisão, mas de acordo com as regras que a determina; (iii) dever é a necessidade de uma ação executada por respeito à lei moral. Nossa última tarefa é pensar se a visão de Kant é plausível. Então, comecemos por pensar se a sua aversão ao consequencialismo traz resultados plausíveis ou não.

Como dissemos, as consequências de uma ação do ponto de vista kantiano não determinam deveres e não são razões suficientes para que não sigamos uma regra moral que passou pelo I.C. Uma estratégia argumentativa comum é checar se uma posição tão forte não soa implausível dado alguma situação específica (essa é uma forma de *reductio ad absurdum*). Então, vamos imaginar uma situação na qual parece que a teoria kantiana daria implicações implausíveis. Pensem, por exemplo, em um mundo possível no qual o presidente dos Estados Unidos sabia que os aviões sequestrados no terrível episódio de 11 de Setembro de 2001 iriam se chocar contra o World Trade Center e o Pentágono. Imagine também que ele soubesse disso com tempo suficiente entre o sequestro e o possível choque das aeronaves contra os prédios. Imagine também que ele deve tomar uma decisão dado duas únicas possibilidades: (a) Não fazer nada e deixar os aviões se chocarem contra os prédios, visto que ele pensa que está proibido de matar pessoas inocentes dentro dos aviões ou; (b) Mandar sua aeronáutica abater os aviões antes deles se chocarem contra os prédios, visto que ele acredita que caso o choque ocorra, tanto mal ocorreria (não apenas a morte das pessoas nos prédios, mas os impactos psicológicos nos sobreviventes e nos familiares das vítimas), que justificaria o abate das aeronaves. Qual as duas ações o presidente deveria tomar neste caso hipotético? Caso a sua intuição aponte para a resposta (a), suas crenças morais são deontológicas kantianas. Aqueles que defendem (a) pensam que os indivíduos são valiosos por si mesmos e que o presidente seria moralmente culpado de assassinato de pessoas inocentes, ou seja, ele quebraria a regra “não devemos matar pessoas inocentes”. Ao passo que se as suas intuições morais apontam para a solução (b), elas são essencialmente consequencialistas. Isso quer dizer

que o ato de abater as aeronaves produziria menos danos e, portanto, um cômputo maior de felicidade/bem-estar. Mas o ponto é: a solução (a) é plausível? Quer dizer deveríamos mesmo seguir cegamente uma regra moral, mesmo que as consequências advindas disso fossem milhares de mortes e centenas de milhares de pessoas sofrendo psicologicamente? Se te parece absurda a aceitação de (a), então a posição de Kant foi reduzida ao absurdo e não deveria ser aceita.

Um segundo problema da teoria kantiana é sobre quais seres compõem a comunidade moral. Como vimos, apenas seres dotados de racionalidade e autonomia, ou seja, seres com dignidade merecem respeito. Todavia, animais não humanos não são racionais como nós e muito menos autônomos como nós, então, estrito senso, nós não devemos nada a eles, dado a teoria de Kant. Mas, para sermos justos, apesar de Kant defender que não devemos nada aos animais não humanos diretamente, nós não deveríamos ser cruéis com eles. O problema está na justificção kantiana para esse dever indireto. Segundo Kant, ser cruéis com animais não humanos é ruim, não porque o sofrimento deles é mau, mas porque se um ser humano for cruel com os animais, isso pode levá-lo a ser cruel para com outros seres humanos. Sendo assim, a ideia kantiana é que a crueldade para animais não humanos cultivaria uma disposição nos seres humanos de serem cruéis com os outros seres humanos. Agora reparem que, não há qualquer defesa de Kant que nos proíba moralmente de utilizar animais como objetos, não há nada de moralmente errado em cultivar animais para alimentação, ou ter animais de estimação para preencher um vazio sentimental nos seres humanos. Adicionalmente, não é errado dispor da vida deles da forma que nos apetece, a não ser se isto nos tornar cruéis para com outros seres humanos. O pensamento kantiano é, em suma, antropocêntrico, o que quer dizer que qualquer consideração que tivermos para com os animais é justificada em função de outros seres humanos e não justificada em consideração aos animais neles mesmos. A maior parte dos pensadores e pensadoras que defendem que os animais não humanos devem ser passíveis de consideração moral pensam que o deontologismo kantiano é uma aberração moral, pois não parece ser plausível pensar que animais não tenham interesses, apesar de não serem racionais como nós. Obviamente que um animal parece ter um interesse em não sofrer, ele não precisa de capacidades altamente desenvolvidas para ser reconhecido moralmente. Adicionalmente, nós reconhecemos moralmente seres

humanos que não possuem a capacidade de autonomia e racionalidade (nem em potência). Por exemplo, nós reconhecemos que devemos respeitar seres humanos senis que não possuem mais a capacidade de racionalidade ou autonomia. Então, porque não estender a nossa comunidade moral para animais não humanos também? Afinal a racionalidade e a autonomia não parecem ser ponto central para alguém integrar nossa comunidade, pelo menos, a nível intuitivo. Apesar das dificuldades com o deontologismo kantiano, há muitos pensadores e pensadoras contemporâneas que defendem algum tipo de deontologismo, entre elas e eles se destacam Onora O'Neill (1990), Christine Korsgaard (1996), Allen Wood (1998), Bernard Williams (1973), Tom Regan (2004) (que defende uma expansão da teoria deontológica a incluir animais não humanos) etc.

Bem, apresentamos nesta seção em linhas bem gerais as principais teorias normativas disponíveis no mercado de ideias filosófico. Mas, ainda nos falta falar algo a respeito da ética aplicada. Grosso modo a ética aplicada se interessa em oferecer respostas aos problemas práticos, utilizando-se dos modelos e dos argumentos desenvolvidos em ética normativa. Ao que segue, faremos considerações de algumas áreas da ética aplicada e seus problemas.

5. Considerações sobre a Ética Aplicada

A Ética Aplicada é uma vastíssima área em crescimento. Isso se dá porque os problemas morais advêm da realidade humana, ou seja, da nossa interação com outros seres humanos, com animais não humanos, com a natureza, com a tecnologia etc., mediados por técnicas e tecnologias criadas por nós. Problemas éticos surgem da realidade, dadas tais interações complexas. Quando chamamos de ética aplicada, muitas vezes damos a entender que vamos apenas aplicar nossos modelos em problemas advindos das atividades descritas acima. Porém, em larga medida, tais problemas nos obrigam a revisar nossos modelos normativos e nossas concepções metaéticas. Por exemplo, ao surgir uma nova tecnologia ou um novo enfoque de nossa atenção moral, devemos-nos perguntar se há parâmetro normativo suficiente para resolver o problema moral que ela levanta; se não há, devemos rever os modelos normativos. Se consideramos que os animais não humanos devem ser incluídos na comunidade moral, devemos saber se nossa axiologia suporta tal ideia; caso

não, devemos modificá-la para abranger os animais, se for essa a conclusão que chegarmos em ética aplicada. O que queremos dizer é que, em ética aplicada, nem sempre é o caso de aplicar modelos pré-estabelecidos em problemas que surgem da realidade, mas muitas vezes é o oposto: dado um problema prático, temos de rever os modelos e seus pressupostos. Sendo assim, sempre é uma via de mão dupla: sim, em partes, em ética aplicada, usamos os modelos pré-estabelecidos para resolver problemas, mas muitas vezes os próprios problemas nos obrigam a revisar os modelos. E a importância desta área não para por aí. Muito do que é desenvolvido em ética aplicada torna-se instrumentos no mundo real, em outros termos, os modelos propostos, os pressupostos assumidos, os valores estabelecidos, as regras morais consolidadas etc. pelas teorias abstratas tornam-se parte de código de condutas profissionais, de legislações de países, de código da comunidade científica etc. Também é digno de nota que alguns problemas estão no entrecruzamento das subáreas da ética aplicada, portanto, não tomem a separação abaixo como estanque. A divisão que apresentaremos tem mais uma finalidade didática do que oferecer uma demarcação clara das áreas. Isso posto, vamos ver algumas subáreas da ética aplicada e seus principais problemas, mas a pretensão aqui não será de discutir as soluções, pois a mera apresentação das principais soluções para cada conjunto de problemas seria demasiado extensa. Então, vamos enunciar apenas os problemas e o que trata cada uma das subáreas eleitas.

(i) A **Bioética**: lida com questões como o início e o fim da vida, com o que as profissionais da saúde devem fazer, como devemos investir os recursos em saúde etc. Sendo assim, são problemas pertencentes a bioética:

- a. Será o aborto moralmente permissível? Se sim, quando? Sempre? Em casos excepcionais (se em casos excepcionais, quais são estes casos)? Se não, em situação alguma?
- b. Será a eutanásia moralmente permissível? Se sim, sob que circunstâncias? Quem deve administrar a eutanásia? Os médicos ou outro profissional?;
- c. Quais os deveres da classe dos profissionais da saúde? Respeitar a autonomia de seus pacientes? Sempre serem beneficentes?

Nunca serem maleficientes?

- d. Deve a classe médica empregar toda a tecnologia disponível para manter alguém vivo, mesmo quando não há mais tratamento curativo?
- e. Qual a forma justa de distribuir os recursos disponíveis para a saúde? Devemos ter apenas um sistema público de saúde? Ou apenas um sistema privado? Ou um misto entre os dois?
- f. Seria permissível utilizarmos a tecnologia médica para melhorarmos os seres humanos, através de manipulação genética?
- g. Seria permissível a clonagem humana?
- h. Seria permissível, se possível, tentar unificar máquinas as pessoas com a finalidade de nos tornarmos mais longevos, ou mais inteligentes, ou mais fortes etc?

Este são apenas alguns dos problemas em bioética, havendo vários outros a serem explorados e, provavelmente, quanto mais as ciências da saúde evoluem, mais problemas dessa atividade aparecerão.

(ii) A **Ética Ambiental**: lida com questões das relações entre os seres humanos e a Natureza de forma geral; São problema da ética ambiental:

- a. Seria permissível utilizarmos os recursos naturais para nos beneficiarmos ou nós devemos algo a natureza?
- b. Seria moralmente permissível controlarmos a natureza de modo a fazê-la funcionar sob o nosso controle?
- c. Se temos que respeitar a Natureza, tal respeito é devido ao quê? Às gerações futuras? Ao próprio ordenamento da natureza? A vida de modo geral?
- d. Será que quem polui mais a natureza, possui um dever de justiça maior de recuperá-la?
- e. Será que devemos viver causando o mínimo impacto na natureza? Se sim, por quais razões? Etc.

Novamente, este é apenas um recorte sobre os problemas que se lida em ética ambiental.

(iii) A **Ética Animal**: lida com os problemas das relações entre animais não humanos e animais humanos;

- a. Será que os animais não humanos possuem direitos como os animais humanos?
- b. É moralmente permissível utilizar animais para o benefício humano?
- c. Seria moralmente permissível utilizar animais para experimentos científicos mesmo que isto cause a eles um grande sofrimento e um baixo ganho para os seres humanos?
- d. Será que temos deveres diretos para com os animais?
- e. Serão os animais moralmente iguais aos seres humanos?
- f. Deveríamos proteger os animais do sofrimento despropositado causado naturalmente, controlando a natureza na qual vivem?

Nem é necessário dizer que este é apenas um recorte dos problemas da área e como se pode ver, há questões se que entrecruzando com outras subáreas da ética aplicada, além de haver várias questões de natureza normativa e metaética.

(iv) A **Ética das relações econômicas**: lida com os problemas das relações econômicas humanas;

- a. Será que tudo está à venda como, por exemplo, órgãos humanos, o sexo, a privacidade etc.?
- b. Qual é o dever das empresas para com os seus funcionários?
- c. O que os funcionários devem as empresas?
- d. Quais os deveres que as empresas têm com relação ao meio ambiente e à sociedade como um todo?
- e. Qual sistema econômico é justo?

- f. Deveria a economia se preocupar em como os recursos materiais são distribuídos na sociedade?
- g. Qual o papel que as empresas deveriam ter no processo político? Se é que devem ter algum?
- h. Seria moralmente correto que empresas patrocinassem campanhas de candidatos a governos?
- i. Como podemos atribuir responsabilidade moral a empresa já que elas não são indivíduos?
- j. Quais os direitos que organizações econômicas devem ter?

Muitos dos problemas em ética das relações econômicas estão ligados aos problemas da ética ambiental, misturados com problemas de filosofia política e do direito, além de questões metaéticas mais gerais como, por exemplo, como devemos entender o que é um agente moral.

(v) A **Infoética**: esta é uma área relativamente nova, mas que estuda as relações entre seres humanos e máquinas (ou inteligências artificiais) e os problemas advindo do emprego da alta tecnologia.

- a. Será permissível deixar uma inteligência artificial comandar aspectos essenciais da vida humana?
- b. Será moralmente permissível utilizar os dados privados dos indivíduos, colhidos por inteligências artificiais fracas, para alimentar algoritmos com a finalidade de vender propaganda?
- c. Será moralmente permissível criar uma inteligência artificial plena?
- d. Será moralmente permissível desenvolver sistemas de interface cérebro-máquina?
- e. Qual os deveres daqueles que desenvolvem alta tecnologia de informação, como, por exemplo, *learning machines*?
- f. Quais os direitos dos indivíduos que usufruem de tecnologias como *learning machines* e algoritmos de captura de informações?
- g. A internet deveria ser limitada por um código de ética? Se sim,

qual código? Se não, por que ela deveria ser totalmente livre de restrições morais?

- h. Devemos fazer tudo que podemos para desenvolver alta tecnologia ou algumas tecnologias nunca deveriam ser desenvolvidas?

Como se pode notar há entrecruzamentos entre as subáreas da ética aplicada, da filosofia política e do direito, além de questões metaéticas e normativas. Mas, importante notar que a existência ou possível existência de tecnologias específicas é que levantam estas questões.

Como dissemos, não pretendíamos oferecer uma lista exaustiva das subáreas da ética aplica. Há várias outras subáreas, como a neuroética, a ética das pesquisas científicas, a ética das relações de gênero e raça, as éticas profissionais etc. etc. A ideia central foi mostrar de modo didático o que está em jogo na ética aplicada. Esperamos ter alcançado tal objetivo.

6. Conclusão

Nosso objetivo com este capítulo foi tentar apresentar, mesmo que de modo breve, a grande área da Ética como disciplina filosófica. Tentamos apresentar de modo satisfatório os principais problemas, as principais posições filosóficas, as razões e argumentos para sustentação das variadas teorias e, principalmente, quais as questões que estão em jogo em cada subárea da disciplina de Ética. Nos desculpamos com o leitor se a exposição aqui sofre de alguma imprecisão, ou se parece por vezes incompleta ou mesmo contenha algum erro. Mas, exortamos ao leitor que tome a nossa exposição como ponto inicial de um estudo que deve ser aprofundado na literatura relevante. Por fim, esperamos que seja de alguma ajuda ao leitor iniciante o que foi dito aqui.

Questões para revisão

1. Qual as características da Ética?
2. Quais as teses centrais de cada modelo teórico que tenta explicar o que é o bem-estar?
3. O que é o princípio utilitarista? Quais as características que ele possui?

4. O que é egoísmo ético?
5. Do ponto de vista do deontologismo kantiano, quais são os elementos relevantes para definir uma ação como moralmente correta?
6. O que é uma Ética das Virtudes?
7. Quais os problemas que se ocupa a Bioética?
8. Quais as relações entre os problemas em ética aplicada?

Questões para discussão

1. Seria possível reduzir todas as descrições sobre o bem-estar à teoria hedonista?
2. É possível formular uma teoria tanto com traços consequencialistas como deontológicos?
3. Será o egoísmo ético uma visão plausível acerca da moralidade?
4. Como poderíamos resolver o problema moral do aborto utilizando uma teoria utilitarista?
5. Do ponto de vista deontológico kantiano, seria moralmente permissível vender órgãos humanos?
6. Do ponto de vista do egoísmo ético, seria permissível vender sexo?
7. Escolha um problema daqueles listados no tópico Considerações sobre Ética Aplicada e tente resolver com um dos modelos normativos.

Sugestões de leitura

A bibliografia em português é um tanto quanto escassa sobre alguns tópicos. As nossas indicações envolvem tanto textos introdutórios como clássicos. Esperamos que os leitores iniciantes na área comecem com as leituras introdutórias e passem as clássicas.

Introdução ao que é Ética

RACHELS, J. & RACHELS, S. Elementos de Filosofia Moral. Porto Alegre: AMGH Editora, 2013. Neste livro, o leitor encontra, no Capítulo 1, uma breve discussão sobre o que é a moralidade. Rachels & Rachels escrevem um livro introdutório e palatável aos iniciantes.

ARISTÓTELES. *Ética a Nicômaco*. 2016. Em *Ética a Nicômaco*, o texto mais clássico de uma *Ética das Virtudes*, Aristóteles discute do que se trata a moralidade.

KANT, I. *Fundamentação da Metafísica dos Costumes*. [Trad. G. A. de Almeida: *Grundlegung zur Metaphysik der Sitten*]. São Paulo: Discurso Editorial/Barcarolla, 2009 [1785]. No texto clássico que funda o deontologismo, Immanuel Kant discute o que é *Ética* já em suas primeiras páginas.

Neste tópico, sugere-se uma leitura dos clássicos e a comparação das diferentes definições oferecidas.

Teorias sobre o Bem-Estar

A bibliografia sobre a natureza do bem-estar é muito escassa em português. Mas podemos indicar, pelo menos duas referências em português e uma em inglês

MULGAN, T. *Utilitarismo*. Petrópolis: Editora Vozes, 2012. Uma referência introdutória em português acerca da concepção de “bem-estar” dentro de teorias utilitaristas pode ser encontrada nesse livro. A discussão acerca do bem-estar se desenvolve no capítulo 4. Mulgan nos oferece uma visão geral das teorias sobre o bem-estar que o utilitarismo incorporou.

DONNER, W. & FUMERTON, R. *John Stuart Mill*. Lisboa: Edições 70. O livro de Donner e Fumerton trata de uma introdução ao pensamento de John Stuart Mill. Porém Donner discute com algum pormenor a tese hedonista de Mill no capítulo dois do livro. Tal capítulo é uma ótima introdução para quem quer começar a estudar a teoria hedonista milliana.

FLETCHER, G. *The Philosophy of Well-Being: An Introduction*. Londres: Routledge, 2016. Talvez a melhor introdução sobre o bem-estar que temos disponível no mercado. O livro de Fletcher introduz de modo claro e abrangente

todas as principais teorias acerca do bem-estar. Cada capítulo é destinado a uma das teorias. O livro contém oito capítulos e discute, por exemplo, o hedonismo, a teoria da satisfação das preferências, teoria da lista objetiva, perfeccionismo, felicidade, teorias híbridas do bem-estar etc. Infelizmente, o livro não se encontra traduzido para o português.

Introdução a Metaética

Como a bibliografia sobre o bem-estar, a bibliografia disponível em português sobre metaética também é escassa. Mas indicamos,

SILVA, M. Moore e os intuicionistas contra o naturalismo. London: *Critica*, 2006. Disponível em https://criticanarede.com/eti_aqa.html. Nesse artigo, aborda-se a contenda filosófica entre intuicionistas (e o que eles defendem) e os naturalistas (e o que eles defendem). O leitor pode entender as bases destas duas posições e porque elas estão em choque no que diz respeito a metaética.

MILLER, A. *Contemporary Metaethics An Introduction*. Cambridge: Polity Books, 2013. Em inglês, por enquanto, este deve ser um dos livros mais acessíveis e introdutórios sobre as posições em metaética analítica. Um capítulo da obra traduzido pode ser encontrado no sítio <https://criticanarede.com/quase-realismo.html>, este capítulo é sobre o quase-realismo de Blackburn.

Introdução à Ética Normativa

RACHELS, J. & RACHELS, S. *Elementos de Filosofia Moral*. Porto Alegre: AMGH Editora, 2013. Como dissemos, este deve ser o melhor livro de nível introdutório sobre ética disponível em português. No capítulo 5, o leitor encontrará uma discussão clara e acessível sobre o Egoísmo Ético. Já nos capítulos 7 e 8 há uma discussão breve, mas esclarecedora sobre a posição utilitarista. No Capítulo 9 e 10, há uma discussão sobre o deontologismo e suas características.

Introdução à Ética Aplicada

RACHELS, J. & RACHELS, S. *A Coisa Certa a se Fazer. Leituras Básicas sobre Filosofia Moral*. Porto Alegre: AMGH Editora, 2014. Nesse livro o leitor encontrará um compilado de textos clássicos sobre ética normativa e ética aplicada. A partir do capítulo 10, o leitor encontrará textos sobre aborto, direito dos animais, fome, pena de morte etc. Há uma ampla gama de posições e problemas de ética aplicada. Apesar de não ser introdutório, os artigos são acessíveis aos leitores familiarizados com os modelos normativos clássicos.

Referências

- ADAMS, R. *Finite and Infinite Goods*. New York: Oxford University Press, 1999.
- ANDERSON, E. "What is the Point of Equality?" in *Ethics*, Vol. 109, No. 2 (Jan., 1999), pp. 287-337
- ANSCOMBE, G.E.M. *Modern Moral Philosophy*. In: *Philosophy*, nº33, p.1-16, 1958.
- ARNESON, R. *Human Flourishing Versus Desire Satisfaction*. *Social Philosophy & Policy Foundation*. 1999
- ARISTÓTELES. *Ética a Nicômaco*. 2016.
- AYER, A. J. *Language, Truth, and Logic*. Nova York: Penguin Books, 2001
- BENTHAM, J. *Princípios da moral e da legislação*. In: *Os Pensadores*. São Paulo: Abril Cultura, 1979.
- BLACKBURN, S. *Essays in Quasi-Realism*. Oxford: Oxford University Press, 1ª Edição (17 junho 1993)
- BRADLEY, F.H. *Ethical Studies*, 2nd. London: Oxford University Press, 1962.
- BRANDT, R. *Morality, Utilitarianism and Rights*. New York: Cambridge University Press, 1992
- _____. *A Theory of the Good and the Right*. Oxford: Clarendon Press, 1979.
- BRINK, D. 'Mill's Deliberative Utilitarianism', *Philosophy and Public Affairs*, 21, 1992, p.67–103.
- _____. *Moral Realism and the foundations of Ethics*. Nova York: Cambridge University Press, 1989.
- CRISP, R. Well-Being, In: *The Stanford Encyclopedia of Philosophy* (Winter 2014

- Edition), Edward N. Zalta (ed.), Disponível em <<http://plato.stanford.edu/archives/win2014/entries/well-being/>>.
- _____. *Reasons and the Good*. Oxford: Clarendon Press, 2006a.
- _____. *Routledge Philosophy Guidebook to Mill on Utilitarianism*. Londres: Routledge, 2006b.
- _____. A distinção fato/valor. Disponível em https://criticanarede.com/eti_factovalor.html, 2006c
- DARWALL, S. "The Second Person Standpoint", Harvard University Press, 2009.
- DONNER, W. & FUMERTON, R. *John Stuart Mill*. Lisboa: Edições 70, 2011
- DWORKIN, R. *Levando dos Direitos a Sério*. São Paulo: Martins Fontes, 2002, p. 357-367.
- D'AVERSA, R. Moral Arguments and the Frege-Geach Problem. PhD thesis, University of York, 2018. Disponível em <http://etheses.whiterose.ac.uk/21981/>
- EPICURO. *Carta sobre a Felicidade (a Meneceu)*. São Paulo: Editora UNESP, 1999.
- DORSEY, D. Moral Distinctiveness and Moral Inquiry in *Ethics* 126 (April 2016): 747–773
- FERKANY, M. The Objectivity of Well Being. *Pacific Philosophical Quarterly* 93, p.472–492, 2012.
- FLETCHER, G. A Fresh Start for the Objective-List Theory of Well-Being. *Utilitas*, Vol.25, p.206–220, 2013.
- _____. *The Philosophy of Well-being. An Introduction*. New York: Routledge, 2016.
- FINNS, J. *Natural Law and Natural Rights*. Oxford: Clarendon Press, 1980
- FOOT, P., *Virtues and Vices*, Oxford: Blackwell, 1978.
- GAUTHIER, D. "Morals by Agreement", Oxford University Press, 1986
- GIBBARD, A. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: OUP Oxford, 1993 (first edition, 1990)
- GRIFFIN, J. *Well-Being. Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press, 1986.
- GEUSS, R. *Philosophy and Real Politics*. Princeton University Press, 2008
- HARTMAN, R.S. The Definition of Good: Moore's Axiomatic of the Science of Ethics. *In Proceedings of the Aristotelian Society New Series*, Vol. 65 (1964 - 1965)

- HOOBES, T. *Leviatã*. São Paulo: EDIPRO, 2015.
- HOOKER, B. *Ideal Code, Real World*. Oxford: Oxford University Press, 2000.
- HURKA, T. *Perfectionism*. Oxford: Clarendon Press, 1993.
- KANT, I. *Fundamentação da Metafísica dos Costumes*. [Trad. G. A. de Almeida: *Grundlegung zur Metaphysik der Sitten*]. São Paulo: Discurso Editorial/Barcarolla, 2009 [1785]
- KORSGAARD, C. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.
- LAUINGER, W. A. The Strong-Tie Requirement and Objective-List Theories of Well-Being. *Ethical Theory and Moral Practice*, p.953–968, 2013.
- LYONS, D. *Forms and Limits of Utilitarianism*. Oxford: Oxford University Press, 1965.
- MACINTYRE, A. *After Virtue*, London: Duckworth, 1985 (2nd Edition)
- MACKIE, J. *Ethics: Inventing Right and Wrong*. New York: Penguin, 1977
- MCDOWELL, J. *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press, 1998.
- MARTINS, R. (ed.). *Classics in Moral and Political Philosophy: An Open Collection*, University of Kansas Libraries, 2018. <http://hdl.handle.net/1808/25944>
- MARTINS, R.(et.al.) (eds.) *Women Throughout History: An Open Collection*, University of Kansas Libraries, 2020. <http://hdl.handle.net/1808/30067>
- MILL, J.S. *Utilitarismo/Liberdade*. São Paulo: Martins Fontes, 2000.
- MILLER, A. *Contemporary Metaethics An Introduction*. Cambridge: Polity Books, 2013.
- _____. Emotivism and the Verification Principle *In: Proceedings of the Aristotelian Society New Series*, Vol. 98, pp. 103-12, 1998.
- MOORE, G.E. *Principia Ethica*. Lisboa: Fundação Calouste Gulbekian, 1999.
- _____. “Principia Ethica” 2nd ed. Cambridge University Press, 1994
- MULGAN, T. *Utilitarismo*. Petrópolis: Editora Vozes, 2012.
- MURPHY, M. *Natural Law and Practical Rationality*. Nova York: Cambridge University Press, 2001.
- NAGEL, T. *The Limits of Objectivity*. Tanner Lectures on Human Values. Oxford: Oxford University, 1979.
- NOZICK, R. *Anarchy, State, and Utopian*. New York: Basic Books, 1974. (disponível também em português: *Anarquia, Estado e Utopia*. Lisboa:

- Edições 70, 2009).
- NUSSBAUM, M. Mill entre Bentham e Aristóteles. *FUNDAMENTO – Revista de Pesquisa em Filosofia*, n. 4, jan–jun – 2012.
- O'NEILL, O. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press, 1990
- PARFIT, D. *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- RACHELS, J. & RACHELS, S. *A Coisa Certa a se Fazer. Leituras Básicas sobre Filosofia Moral*. Porto Alegre: AMGH Editora, 2014.
- _____. *Elementos de Filosofia Moral*. Porto Alegre: AMGH Editora, 2013.
- RAND, A. *The Virtue of Selfishness*. Nova York: Penguin Group, 1964.
- RAILTON, P.; DARWALL, S.; & GIBBARD, A. *Metaética Algumas Tendências*. Florianópolis: EDUFSC, 2013
- RAWLS, J. *A Theory of Justice*, Harvard University Press, 1999
- REGAN, T. *The Case For Animal Rights*. São Francisco: niversity of California Press, 2004.
- RICE, C. Defending the Objective List of Well-Being. *Ratio*, XXVI, p. 196–211, 2013.
- SANTOS, B.A.G *Utilitarismo e Direitos Morais Básicos*. Tese de Doutorado, 2017, Disponível em <https://repositorio.ufsc.br/handle/123456789/177775>.
- _____. Três Teorias sobre o Bem-estar. In. *Crítica*. London, 2017. Disponível em https://criticanarede.com/fil_felicidade.html
- _____. Objective List Theory of Well-Being as an Explanatory Theory. In: *Ethic@*, v.14, n 3, 2015. Disponível em <https://periodicos.ufsc.br/index.php/ethic/article/view/41550>
- SANTOS, B.A.G & MERLUSSI, P. Resenha de DARWALL, S.; GIBBARD, A.; RAILTON, P. *Metaética: algumas tendências*. (Org) Darlei Dall'Agnol; tradução de Janine Sattler. Florianópolis: Editora UFSC, 2013.
- SARCH, A. Multi-Component Theories of Well-Being and Their Struture. *Pacific Philosophical Quarterly* 93, p.439-471, 2012
- SEN, A. “O que nós queremos de uma Teoria da Justiça?” *Journal of Philosophy* 103: 215-38, 2006,
- SMART, J. J. C. & WILLIAMS, B. *Utilitarianism: for and against*. Cambridge: Cambridge University Press, 1973.
- SINGER, P. *Ética Prática*. São Paulo: Martins Fontes, 2012.
- _____. *Animal Liberation*. New York: Ecco Press, 2001.

- _____. *Ethics*. Oxford: Oxford University Press, 1994.
- SINNOTT-ARMSTRONG, Walter, "Consequentialism", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.)
Disponível em
<https://plato.stanford.edu/archives/sum2019/entries/consequentialism/>.
- SKORUPSKI, J. The Place of Utilitarianism in Mill's Philosophy. In: WEST, H. (org) *The Blackwell Guide to Mill's Utilitarianism*. Oxford: Blackwell Publishing, 2006
- SKORUPSKI, J. *Why read Mill today?* New York: Routledge, 2006.
- SWANTON, C. *Virtue Ethics: A Pluralistic View*, Oxford: Oxford University Press, 2003.
- VALERIUS, J. Autonomy, Subject-relativity, and Subjective and Objective Theories of Well-being in Bioethics. *Theoretical Medicine* 24: p.363–379, 2003.
- WILLIAMS, B. *Utilitarianism: For and Against*, Cambridge University Press, 2008
- _____. *In the Beginning Was the Deed*. Princeton: Princeton University Press, 2008.
- WOOD, A. "Kant on Duties Regarding Nonrational Nature". *Proceedings of the Aristotelian Society Supplement*, LXXII: 189-210, 1998.
- ZAGZEBSKI, L. *Virtues of the Mind*, New York: Cambridge University Press, 1996.

5

Estética

André Luiz Alves Pereira

1. Introdução

Você já deve, em algum momento, ter-se deparado com alguma questão que é tradicionalmente investigada pela área da filosofia a qual damos o nome de Estética. Tente se lembrar de uma belíssima música e do modo como você se sente ao ouvi-la com atenção. Assim que você a ouve é provável que se sinta de um modo diferente do habitual. Dentro desse estado, pode já ter-se perguntado “o que faz esta canção *ser* tão bela?”, ou, de maneira ligeiramente diferente, “por que esta canção me faz *sentir* algo tão diferente?”. Caso essas perguntas nunca lhes tenham passado pela cabeça, pense um pouco nelas.

Você pode também ter visto uma paisagem que tenha produzido uma experiência parecida com a audição daquela canção. Depois de um memorável pôr do sol ou a visita a alguma paisagem belíssima, como um alto de morro ou uma praia paradisíaca de um azul cristalino e ondas redondinhas, você volta para casa com a impressão que teve essa experiência antes. Pode ter associado alguma coisa da bela paisagem com a bela canção. Pode ter passado pela sua cabeça (embora num lapso de tempo): ‘o que este pôr do sol e aquela canção possuem em comum, já que me sinto de modo tão semelhante diante dessas duas coisas?’.

Costumamos dizer que uma enorme diversidade de coisas é bonita. Alguns reparam na combinação de um pequeno jardim muito colorido, com uma casa ao fundo; outros no formato arrojado e na capa chamativa de um livro de ficção científica; num prédio público, com suas sóbrias colunas revestidas de gesso ou seu formato retangular cheio de aço e vidro cristalino escuro; numa igreja enorme, com seus contrastes de luz marcantes, ou uma pequena, com uma clareza mais discreta e singela; um automóvel, com suas formas

arredondadas, tendendo ao chão, ou agressivas e retangulares, que parecem projetá-lo para frente. Diante de objetos assim, não parece estranho que um amigo vire para o outro e diga “minha nossa: isto é uma verdadeira obra de arte, não é mesmo?”, para expressar o quanto essas coisas são diferentes das demais de algum modo especial.

Os mais curiosos podem ter notado que uma série de objetos e ações recebem com mais frequência essa atenção especial. Pense na escultura chamada *Davi*, feita por Michelangelo no século XVI, localizada atualmente em Florença, na Itália, que recebe milhares de turistas todos os dias, fora as pesquisas virtuais de sua imagem. Considere uma pintura como a *Guernica*, uma pintura em tela feita por Pablo Picasso em 1937, atualmente sob responsabilidade de um museu em Madrid, alvo de batalhas políticas desde o momento de sua primeira exposição pública – fora a grande reverência que recebeu por um número enorme de pintores desde então. Lembre-se de romances celebrados como verdadeiros “clássicos” por estudiosos da literatura, como *O Processo* de Franz Kafka ou outros altamente celebrados pelo público – e nem tanto pelos estudiosos – como a série de romances do mundialmente famoso Harry Potter, de J. K. Rowling. Esses exemplos bastam: todos eles recebem uma atenção e um valor muito diferentes do que uma bela cafeteira ou uma bela televisão – embora muitos dos seus admiradores se queixem há muito de que deveriam ser mais apreciados, pois a indiferença ou ausência de preparo para essas experiências colocariam valiosas experiências humanas em risco.

Existem instituições – como os teatros, os museus e as galerias de arte – que cuidam desses objetos e os organizam para que todos os dias muita gente possa experimentá-los com a atenção que não costumamos dar aos postes de luz nas ruas da cidade. Boa parte desses objetos especiais estão localizados em instituições assim, mas muitos, hoje talvez a maioria, não estão. Pensem em romances ou filmes. Onde eles estarão?

Mesmo obras que usualmente são guardadas e expostas por galerias e museus podem ser deslocadas para outros lugares. Alguns artistas, já faz mais de século, insatisfeitos com as limitações que museus e galerias impunham à produção artística (e com as expectativas tradicionais do público), pensam que o melhor seria produzir objetos para serem apreciados em outros locais, colocando-os mais próximos do cotidiano das pessoas, como em grandes muros ou no centro de uma praça por onde passe muita gente todos os dias. Ou

simplesmente produzindo canções para tocar nos nossos *streamings* de música.

Muitos já devem ter notado que a maior parte das pessoas se refere a todos esses objetos e ações especiais, feitos por seres humanos, utilizando o termo *arte*. E não apenas para expressar o quanto apreciam tais objetos, mas simplesmente porque todo mundo os classifica dessa maneira. Isso faz com que imediatamente nos lembremos deles. É uma palavra e um significado corrente em muitos idiomas. Se você parasse alguém na rua e perguntasse se a *Mona Lisa* é um exemplar de *arte*, muitos poderiam desconfiar da sua sanidade ou das suas intenções. Outros, no entanto, poderiam entender a questão de modo mais geral: afinal, o que torna *Mona Lisa* uma obra de arte? E indo além: o que torna qualquer objeto uma obra de arte?

Antes nos perguntávamos sobre a beleza de certos objetos e o que eles tinham em comum; agora nos perguntamos por que objetos de certo tipo recebem uma classificação que aparentemente lhes dá tanta importância. Mas alguém poderia se lembrar que muitas obras de arte – chamemo-las assim por enquanto – não são tão apreciadas quanto outras. Um conto escrito por um estudante do ensino médio pode ser muito bom para mostrar como ele vem desenvolvendo certas capacidades, como a criatividade, a imaginação e a escrita, mas ninguém parece exigir que ele faça um conto tão bom quanto um de Machado de Assis, de Virgínia Wolf ou de Jorge Luís Borges. Por uma razão simples: parece que nos contos desses escritores clássicos há certas qualidades não tão fáceis de serem alcançadas que os tornam *melhores* do que os resultados de um exercício escolar. Que qualidades são essas? Por que pensamos que há um valor geral maior das qualidades de obras como as de Borges do que a de um exercício de um estudante? Aliás, há algum valor geral que podemos utilizar para diferenciar esses dois objetos escritos? Note que não estamos negando que o que o estudante escreveu é um conto e pode contar como arte; a questão é saber seu valor em relação aos demais contos. Disso parece se seguir que há duas coisas distintas (e talvez relacionadas) as quais devemos dar maior atenção: uma é a de saber o que torna um objeto um exemplar de *arte*; outra é saber qual é o valor que cada uma dessas obras tem. Indo além, podemos nos perguntar que valor a arte tem *em geral*.

Todas as perguntas que fizemos até agora, a partir das quais poderíamos nos fazer muitas outras, quando persistem nas comunicações entre as pessoas, são indícios de que existe algum interesse em encontrar uma boa

resposta para elas. Respostas que, se não forem completas, ao menos afastam algum mal-entendido que não estejamos percebendo. De maneira bem concisa, muita gente tentou ir ao cerne dessas questões, perguntando e tentando responder ‘o que é a beleza?’, ‘o que é uma obra de arte?’, ‘por que uma obra parece ter maior valor do que outra?’.

2. Estética

Começamos este texto, lendo a afirmação de que essas perguntas são investigadas por uma área que tradicionalmente tem sido chamada de ‘Estética’. Antes de ser compreendida apropriadamente, no entanto, a afirmação requer um adendo. ‘Estética’ é um termo que variou bastante ao longo da história e, mais especificamente, na história da filosofia. Ainda hoje essa palavra pode assumir diferentes significados. Neste texto, vamos nos concentrar em quatro diferentes questões, que já foram mencionadas vagamente. As quatro questões a seguir têm sido consideradas as mais fundamentais dessa área, fundamentais porque são anteriores a várias perguntas e estão implícitas de modo menos simples e direto em vários dos nossos questionamentos e experiências cotidianas. São elas:

1. O que é a arte?
2. O que é a beleza?
3. O que é o gosto?
4. Que valor a arte tem?

Quadro 1: O Nascimento da Estética

A Estética começa sua sinuosa trajetória no primeiro quarto do século XVIII. As mudanças de significado do termo ‘estética’ na obra daquele que é reconhecido como seu primeiro proponente podem ser tomadas como prenúncios das várias mudanças que o termo sofreria a seguir. Em 1735, o jovem Alexander Baumgarten introduziu pela primeira vez o termo em sua dissertação universitária, intitulada “Considerações filosóficas sobre alguns aspectos pertinentes ao poema”. Nesse texto, ‘estética’ buscava designar “uma ciência de como as coisas são conhecidas por meio dos sentidos”. Depois de várias reformas nessa definição feitas em textos posteriores,

Baumgarten chegou a sua formulação definitiva: “Estética (a teoria das artes liberais, a gnoseologia menor, a arte de pensar belamente, a arte do análogo da razão) é a ciência da cognição do sensível”.⁹¹

Baumgarten concordava, como boa parte dos filósofos de sua época, que existe uma separação incontornável entre as representações “internas” que temos do mundo e o mundo tal como ele é. A filosofia, desta forma, teria a tarefa de avaliar em que condições essas representações coincidem com os vários aspectos do mundo e, no fim das contas, no modo como ele é. Dentre os vários modos de os seres humanos representarem o mundo, pensavam eles, alguns dependem de experiências com o “mundo externo”, as quais são mediadas pelos nossos sentidos – a visão e a audição, por exemplo – e outras, referentes exclusivamente ao “mundo interno”, pareciam-lhes não depender dos sentidos. Baumgarten estava especialmente interessado no modo como adquirimos um tipo de conhecimento – que julgava diferente dos demais – exclusivamente por meios dos sentidos e qual seria a natureza desse tipo de conhecimento. Esse tipo de conhecimento seria crucial nas nossas experiências com as belas-artes, utilizando-as largamente em suas reflexões. Dizia que “o conhecimento sensível é o complexo de representações que subsistem abaixo da distinção”. Isso não queria dizer que essas representações sejam *obscuras*, como pensavam vários dos filósofos do seu tempo; mas que requerem um tratamento cuidadoso e especial no que elas têm de específico. Daí sua proposição de que uma “ciência” – um conjunto delimitado de métodos e problemas a serem investigados – deveria surgir para tentar solucionar as questões levantadas por esse tipo de representação.

Muitos filósofos não viram relevância ou sequer deram atenção à proposta de Baumgarten, por vários motivos filosóficos. Outros, no entanto, estavam se dedicando a questões similares às propostas por ele e, à medida que foram tomando conhecimento da distinção de uma possível “ciência estética”, foram vendo confirmadas muitas das suas expectativas e notando que suas investigações não estavam sós, passando a construir, aos poucos, algo como um diálogo entre si.

Essas discussões foram tomando um rumo próprio e, em seu conjunto, formaram, ao menos em parte, o que hoje damos o nome de ‘Estética’. Desde então, vários e vários filósofos foram alterando as perguntas e oferecendo novas respostas quando lhes pareceu conveniente.

⁹¹Baumgarten, apud, Kivy. P. 27

3. O que é a arte?

Pode parecer estranho, mas quando perguntamos o que é a arte já temos uma porção de coisas pressupostas na pergunta. Na verdade, temos uma porção de coisas pressupostas quando perguntamos o que qualquer coisa é, por isso é preciso dar alguma atenção a esses pressupostos. Por exemplo, se você parar e perguntar ‘o que é uma cadeira?’ (exemplo clássico numa aula de filosofia), você provavelmente pode estar esperando alguma característica (ou um conjunto delas) que diferencia as cadeiras de qualquer outra coisa. Quando perguntamos ‘o que é arte?’, parecemos ter a mesmíssima expectativa. Se formos capazes de encontrar as características que diferenciam os objetos artísticos de todos os outros objetos, estaremos muito próximos de responder a questão.

O problema é que quando estamos falando sobre arte, não é raro que haja muito desacordo inicial sobre quais objetos contam como uma “obra de arte”.⁹² Isso coloca uma dificuldade adicional a esse tipo de investigação. No nosso cotidiano utilizamos diversos conceitos corretamente, sem nos vermos na menor necessidade de defini-los. Aprendemos vários deles ainda muito pequenos e os utilizamos pelo resto das nossas vidas. Considere o conceito de “verde”. Embora dificilmente você tenha pensado em oferecer uma definição de “verde”, é muito provável que você já tenha que ter lidado com esse conceito. As várias pessoas que te ensinaram a utilizá-lo provavelmente não precisaram “defini-lo” para você. Tudo que você precisou aprender quase imperceptivelmente foi diferenciar quando essa cor ocorre nos objetos e quando não ocorre. Notando o acordo entre as pessoas nessas atribuições do conceito de verde, talvez tenha se dado por satisfeito.

Essa é uma forma muito difusa de definir os conceitos e talvez esteja presente em todo lugar onde haja seres humanos. Essa é uma forma de definir um conceito que muitos chamam de *ostensiva*. Se alguém pergunta o que é uma cadeira, uma maneira de oferecer uma definição seria apontar para alguns exemplares do que queremos nomear dessa forma. Crianças e aprendizes de idiomas aprendem diversos conceitos assim e utilizam o conceito de maneira bem-sucedida na maior parte das vezes que precisam. Os problemas começam

⁹² Daqui para frente usaremos os termos ‘arte’ e ‘obra de arte’ como sinônimos.

a aparecer quando há a necessidade não apenas de nomear este ou aquele objeto, mas ter um critério que nos permita diferenciar determinado objeto ou qualidade de todos os demais. A arte é um excelente exemplo, onde há mais desacordo sobre quais objetos devem receber este ou aquele termo e, sobretudo, *por que* é que outro tipo de definição ganha interesse.

Quais objetos devemos incluir “ostensivamente” no conceito de arte logo de saída? Muitos não disputariam que a *Mosa Lisa*, *Davi* e a *Guernica* são bem classificados como obras de arte. Mas muitos não estariam dispostos a conceder que outras obras, como a *A Fonte*, feita por Marcel Duchamp, seja incluída na investigação sobre o que é uma obra de arte.

Quadro 2: A Fonte, de Marcel Duchamp

Marcel Duchamp foi um artista francês sumamente polêmico em seu tempo. Tornou-se mundialmente famoso por produzir uma obra que obrigou muita gente a pensar sobre o que, afinal de contas, é uma obra de arte. Essa obra, ele chamou de *A Fonte* e se encontra atualmente no acervo da Indiana University Art Museum, nos Estados Unidos. Podemos descrevê-la como um mictório com algumas inscrições em tinta preta, disposta numa instituição artística – como um museu – para apreciação do público. Exposta em 1917, colocava algumas questões interessantes: uma obra de arte precisa ser produzida pelo artista? Se sim, em que medida? É suficiente para um objeto ser considerado uma obra de arte que um artista tenha uma ideia e selecione um objeto, alterando-o muito pouco ou nada, para exibi-lo numa instituição dedicada à arte? Será que a definição do que é uma obra de arte se deve ao que o artista produz, ou simplesmente ao modo como as pessoas percebem e dão significado ao objeto? Ou uma instituição – como uma galeria e um museu – pode falhar em algum ponto dos seus critérios de seleção do que conta ou não como uma obra de arte – tornando a exposição de *A Fonte* apenas um mal-entendido?

Essa é uma questão que precisa sempre ser levada em conta, pois pode ser que um dos efeitos indesejáveis de uma definição estrita possa ser o de excluir indevidamente a inovação e mesmo a possibilidade de mudança de quais coisas devem ser incluídas no conceito de “obra de arte”. Esse foi um ponto debatido por bastante tempo e, para dizer a verdade, não parece ter se encerrado. Falaremos mais dele ainda quando discutirmos uma teoria que propõe outra forma de oferecer definições.

Pode ser que, deixando um pouco de lado esses casos especiais, seja

mais interessante olhar para o que seguramente podemos contar como uma obra de arte, investigando o que seguramente elas têm em comum. Adiante, talvez, ganhando uma compreensão maior dos casos mais seguros, podemos retornar a casos como o de *A Fonte* mais bem equipados conceitualmente.

Começemos separando algumas características e buscando refletir se elas são *condições necessárias* de uma obra de arte. Se quisermos definir uma cadeira, por exemplo, teremos que levar em conta o fato de que ela é um artefato produzido por seres humanos. Independentemente de qual seja a definição de cadeira, seria estranha uma definição que não levasse essa característica em consideração. Como sabemos que há muitos artefatos produzidos por seres humanos que não são cadeiras, podemos concluir que essa condição não é *suficiente*. Embora tenhamos que levá-la em conta, não poderemos nos dar por satisfeitos se quisermos encontrar a *definição*. Uma definição nada mais é do que uma descrição de quais condições individualmente *necessárias* são conjuntamente *suficientes* para definir um objeto. Diferentemente das definições ostensivas – onde não há necessidade de descrever tais condições –, essa é uma definição “explícita” – que busca descrever quais são as condições necessárias e suficientes que fazer determinada coisa ser o que é.

Quadro 3: Definições explícitas

Definir explicitamente um conceito significa apresentar todas as condições que têm de ser satisfeitas para que esse conceito seja aplicado de maneira bem-sucedida. O procedimento padrão para apresentar definições explícitas é o uso de afirmações ou sentenças bicondicionais, isto é, uma sentença do tipo "A se e somente se B". Uma sentença bicondicional é composta pela conjunção de duas sentenças condicionais: "se A, então B" e "se B, então A". A primeira condicional diz que B é uma condição suficiente para A; a segunda diz que A é uma condição suficiente para B. Juntas elas dizem que A é uma condição necessária e suficiente para B". O que quer dizer intuitivamente que se algo é A, então é B, e se algo é B então é A. Considere então a seguinte definição de cadeira:

X é uma cadeira sse (forma abreviada de 'se e somente se'): (i) é um artefato; (ii) com assento individual; (iii) com encosto.

Assim, espera-se delimitar adequadamente o conjunto de todas as cadeiras e, com tal delimitação, distingui-las de todos os demais objetos.

Mas será que toda definição em termos bicondicionais é boa? Considere a definição acima. Essas condições necessárias incluem algum objeto que não seja uma cadeira? A primeira vista, infelizmente, sim. Um sofá, por exemplo. Um sofá pode ser descrito do mesmo modo, pois (i) é um artefato; (ii) com assento individual; (iii) com encosto. Mesmo que isso não seja suficiente para delimitar o que os sofás são, e alguma condição seja controversa – como a número (ii) –, a definição de cadeira parece incluí-los, o que não deveria ocorrer quando o sinal lógico da bicondicional aparece. O que é muito comum, pois boa parte da atividade filosófica consiste em tentar oferecer definições explícitas e, então, revisá-las, visando uma compreensão maior da questão. Algo que costuma ser considerado uma cadeira fica excluído de acordo com a definição? Aparentemente não. Então, o problema parece estar apenas de um lado da definição explícita.

3.1 Teorias da imitação

Algumas tentativas de definição explícita da arte ficaram conhecidas como teorias da imitação. Essas foram teorias muito influentes ao longo da história da filosofia – tendo aparentemente sido propostas, em versões muito distintas, por Platão e Aristóteles. Muita gente que pensa sobre o assunto busca incorporar algumas das suas qualidades. Podemos dizer que a afirmação central que elas fazem é a seguinte: algo é uma obra de arte sse (i) é produto humano e (ii) imita a aparência de um outro objeto. Que sejam os seres humanos, os produtores das obras de arte, não parece levantar nenhuma controvérsia, mas que elas imitem sempre outros objetos parece requerer um exame mais detido.

Pensemos no caso do *Davi*, de Michelângelo: será que essa obra se encaixa bem no conceito de obra de arte oferecido pelas teorias da imitação? A obra foi certamente produzida por um ser humano – e nesse caso podemos especificar inclusive qual, embora a afirmação central dessa teoria nem nos peça isso. *Davi* se encaixa bem na condição (i). *Davi* imita aparência de um outro objeto? Parece que sim: imita a aparência de um ser humano do sexo masculino. (Note que “objeto” aqui é uma noção bastante ampla. Qualquer coisa no mundo pode ser um objeto, a princípio). Então, parece que se todas as obras de arte forem como *Davi*, a teoria da imitação é bem-sucedida em oferecer uma definição explícita de obra de arte.

Mas para avaliar uma teoria, precisamos tentar encontrar alguma obra que dificilmente alguém negaria ser uma obra de arte e que não se encaixe na

condição (ii). Será que ela existe? Talvez. Considere o quadro *Guernica* de Pablo Picasso. Embora possa haver uma controvérsia aqui e ali, é difícil negar que essa pintura seja um exemplar de obra de arte. *Guernica*, no entanto, não parece imitar a aparência de nenhum outro objeto, ao menos não como estamos acostumados a vê-los por aí. Algum estudioso da arte, no entanto, poderia argumentar que esse quadro visa imitar por meio da sua *forma*, isto é, pela maneira como está organizada sua composição, o sentido humano profundo de dada situação histórica ou, quem sabe, um aspecto importante da “condição humana”. Se esse crítico estiver correto, mesmo que tenhamos que rever um pouco o que entendemos por “imitar”, parece que a teoria da imitação continua bem-sucedida.

Agora tente imaginar uma pintura como as que pintou o bastante conhecido Jackson Pollock. Utilizando uma técnica de pintura inovadora conhecida como *dripping* (gotejamento), Pollock não estava muito interessado nas aparências dos objetos em suas pinturas, nem em representar coisa nenhuma. Não utilizava pincéis nem cavaletes para pintar, pois não necessitava da precisão associada à pintura que busca representar a aparência dos objetos. Deixando com que a tinta “pingasse” mais ou menos ao acaso na tela, ia procurando combinar as cores e traçando formas inusitadas, cujo resultado final deveria provocar a percepção do espectador. Se as obras de Pollock contarem como obras de arte, então há algo errado com a definição das teorias da imitação, pois seus quadros parecem mostrar que imitar um objeto não é uma condição necessária para que algo seja uma obra de arte.

Poderíamos ainda considerar se a definição inclui outras coisas que não temos em vista. Pense numa maquete de um condomínio residencial ou de uma obra pública. Elas parecem se encaixar perfeitamente na definição da teoria da imitação, pois satisfazem as duas condições. Embora algumas maquetes possam ser muitíssimo bem feitas, sendo portadoras, quem sabe, de rara beleza, dificilmente diríamos que elas são obras de arte. Então, parece existir outro problema com a afirmação central da teoria da imitação, por incluir objetos que não são obras de arte.

A teoria da imitação é capaz de se sair bem em alguns casos, mas nem tanto em outros. Se achamos que ela é uma boa teoria e merece ser desenvolvida, precisamos olhar com cuidado para esses problemas.

3.1.1 Aristóteles

O problema de se saber se os filósofos da Antiga Grécia tinham ou não em vista uma definição do conceito de ‘arte’ de modo suficientemente próximo ao que pensamos atualmente se torna saliente nas obras de Aristóteles que chegaram até nós. A mais conhecida a tratar do assunto, a sua *Poética*, é vista pela maior parte dos seus estudiosos como notas para serem utilizadas em suas aulas. Era um apoio à comunicação oral, portanto, e não foi composta com a intenção de circular entre um público amplo. Felizmente o texto foi transmitido por sucessivas gerações e hoje podemos estudá-lo.

Talvez seja essa a explicação de por que Aristóteles não oferece uma definição do que é a *Poética* nesse texto. A posição que ela ocupa no conjunto das investigações do filósofo normalmente envolve levar em conta o conjunto do seu pensamento. Poderemos fazer apenas algumas indicações nesse sentido.

Aristóteles também defende que arte é imitação; mas por “imitação” ele entende algo bem diferente do seu professor Platão. Essa diferença não se deve meramente a uma discordância tópica a respeito da natureza da imitação, mas a diferenças bem mais amplas sobre o que é a própria realidade (embora ‘realidade’ seja um termo parcial aqui). Aristóteles rejeita a teoria das Formas tal como apresentada por Platão; em seu lugar, propôs toda uma nova maneira de entender o mundo. (Veremos os traços gerais da teoria das Formas adiante).

Em linhas gerais, para esse pensador, a poesia poderia ser investigada de acordo com sua própria realidade e por suas próprias regras de funcionamento. Ainda caberia ao filósofo a busca pelas essências universais que definem cada coisa – (arboridade da árvore), por exemplo –, buscando-se pelas “causas” que constituem cada entidade. Os filósofos buscam o universal, a classe ou o gênero à qual as entidades pertencem, sem que essas sejam categorias de classificação meramente “aplicadas” às entidades, e sim partes constitutivas delas e da própria realidade. Nesse caso, os indivíduos – uma mesa qualquer, um lápis qualquer, uma pedra qualquer – não existem sem que façam parte de classes, gêneros e universais.

Essa classificação está relacionada aos “tipos de conhecimento” definidos pelo filósofo, que são inicialmente três: a experiência, a arte (“técnica”) e a ciência. Há ainda um tipo de conhecimento que está acima dessas três, ao qual deu o nome de “filosofia primeira” – aquela que se ocupa dos universais

enquanto tais. A experiência é o tipo mais comum e difuso de conhecimento; “até” os animais seriam capazes de atingi-lo. A “arte” envolveria uma generalização maior do conhecimento das causas e dos processos gerais de como cada coisa vem a ser o que é. A ciência seria aquele conhecimento sobre gêneros como, por exemplo, os seres vivos (biologia), o movimento (física) etc.

Como estamos interessados em nos aproximar um pouco do que pensava Aristóteles sobre a “arte”, avancemos nesse sentido. Em outra célebre definição, o filósofo nos diz que “arte” (no sentido de “técnica”) é “uma disposição racional da capacidade de fazer”, (...) “uma disposição da capacidade de fazer, envolvendo um raciocínio verdadeiro”. Aristóteles divide ainda dois tipos de arte: aquelas que resultam em algum “produto” e aquelas que não resultam. A arte de curar e arte de governar são dois exemplos de artes que não deixam nenhum “fabricado” como resultado. O que importa nas artes é que elas são atividades governadas por regras e que visam a atingir objetivos práticos determinados, descobrindo o que deve ser feito em cada situação e caso particulares. A Poética seria um tipo de “arte” (técnica) que deixaria um produto – uma obra –, capaz de nos mostrar a natureza e a finalidade de cada gênero literário.

Com o que dissemos ficará mais fácil entender o que Aristóteles entende por “Arte mimética” que se aproxima muito mais daquele nosso interesse em saber por que a *Mona Lisa* de Leonardo da Vinci ou o *Rei Lear* de Shakespeare são obras de arte. De acordo com as distinções aristotélicas, a poética faria parte das “artes” (técnica), do subtipo que deixa um produto – uma obra.

A *Poética* começa por distinguir três aspectos da imitação: o meio da imitação, o objeto da imitação e o modo da imitação. Por meio desses aspectos, o filósofo procede sua investigação dos vários gêneros literários, até chegar no gênero que tem em vista: a tragédia. Vale a pena acompanhar a concisa definição que ele oferece:

A tragédia é a imitação de uma acção elevada e completa, dotada de extensão, numa linguagem embelezada por formas diferentes em cada uma das suas partes, que se serve da acção e não da narração e que, por meio da compaixão e do

temor, provoca a purificação de tais paixões.⁹³

Nessa definição aparecem sua teoria da arte – a imitação – e as distinções que traçam os limites de cada gênero literário: o objeto da imitação (a ação elevada), o meio da imitação (a linguagem embelezada, no sentido de aprazível e agradável) e o modo da imitação (a ação e não a narração). A finalidade de tais produções seria a famosa “purificação” (catarse), por meio das paixões do temor e da compaixão.

Há muito a se refletir sobre o pensamento de Aristóteles sobre a arte. Tanto da sua definição acerca da natureza quanto dos efeitos da arte, parece se seguir uma visão positiva do que chamamos de ‘arte’. Se bem realizada, ela pode provocar um efeito muito benéfico – a catarse – e não há nada em sua natureza que nos afaste do mundo tal como ele é. Mas essas conclusões parecem ter como condição necessária uma visão específica do que é a realidade, de modo que podemos concluir, ao menos diante desses dois breves casos, que uma visão sobre o que é a realidade como um todo e o que é a arte estão estreitamente ligadas em seu pensamento, incluindo o espinhoso problema de se saber qual é o valor da arte e que natureza tem esse valor.

A “imitação”, mesmo que compreendida de modo diferente, pode ser uma condição necessária do que entendemos por ‘arte’, mas dificilmente será suficiente. (É bom frisar mais uma vez que há de se levar em conta, como dito antes, que talvez fosse outro o conceito que os gregos utilizavam).

Em Platão, há ligações com a vida política e com a moralidade bastantes salientes, como veremos, ao passo que, em Aristóteles, essas ligações parecem não existir, ou existir de maneira bastante indireta. Nos dois casos, suas respostas sobre o que é a realidade e o seu compromisso com uma concepção de filosofia como uma atividade que busca conhecer o que o mundo como um todo é foram cruciais para responder a essas perguntas. Tais problemas e muitos outros estão longe de encontrar uma palavra definitiva; cabe a você refletir sobre eles.

⁹³ ARISTÓTELES. **Poética**. Trad.: Ana Maria Valente. Lisboa: Fundação Calouste Gulbenkian, 2008. p. 47.

3.2. Teorias da expressão

Uma teoria alternativa à teoria da arte como imitação enfatiza menos as características objetivas das obras de arte – se representam bem o objeto que pretendem representar, por exemplo – e enfatizam mais a relação do produtor com suas obras. Essa família de teorias ficou conhecida como teorias expressivas da arte. Sua afirmação central é a seguinte: algo é uma obra de arte se (i) é produto humano, (ii) por meio do qual o artista expressa estados emocionais e sentimentais. Repare bem que a emoção e o sentimento em questão pertencem ao artista; na definição nada se fala sobre a capacidade da obra de desencadear emoções ou sentimentos, como aparece no pensamento de Aristóteles sobre a tragédia.

Essa visão ganhou bastante força durante o final do século XVIII e início do XIX com a formação do chamado Romantismo. Muitos filósofos puseram uma forte ênfase na ligação entre a subjetividade dos artistas e as obras que eles produziam. Todos pareciam bastante de acordo com a afirmação central de uma teoria da arte como expressão do artista. Muitos dos seus proponentes eram eles mesmos artistas e tentavam teorizar levando em conta suas experiências pessoais de criação. E os testemunhos íntimos de muitos outros artistas conferiam uma grande força de saída e confiança nessa teoria. Afinal, quem melhor que seus próprios *criadores* para dizer o que é a arte?

Esta é uma teoria precisa que nos permite distinguir claramente se um objeto é ou não uma obra de arte. Se expressar sentimentos e emoções por meio de um objeto, é uma obra de arte. Se não expressar, não é. A seu favor, podemos dizer que aspectos da teoria jamais foram abandonados completamente da conversa pública sobre arte. Se um crítico ou um visitante de passagem pelo Museu de Arte Moderna de Nova York desse a sorte de deparar com a famosa tela *O Grito*, de Edvard Munch, poderia muito bem formar a crença de que a tela realmente expressa uma grande *angústia diante da existência*, como muitos já afirmaram. Uma crença do tipo está muito próxima da teoria expressiva da arte, embora não seja a mesma coisa. Isso porque uma teoria expressiva da arte não busca apenas destacar um aspecto importante das obras de arte. Ela visa oferecer uma resposta completa para a pergunta ‘o que é a arte?’.

A teoria encontra uma série de dificuldades. A primeira delas: para se

saber se uma obra como *Davi* e *Guernica* são obras de arte, precisaríamos antes saber se expressam alguma emoção ou sentimento dos seus respectivos realizadores, Michelângelo e Picasso. Temos alguns testemunhos de Picasso sobre sua obra, que permitiriam justificar a nossa crença, por exemplo, de que a tela expressa seus sentimentos sobre o bombardeio alemão à cidade de Guernica, durante a triste e sangrenta Guerra Civil Espanhola. Com Michelângelo, as coisas ficariam mais complicadas, pois temos poucos relatos em primeira pessoa do artista disponíveis. Deveríamos suspender a nossa decisão de considerar ou não *Davi* uma obra de arte porque não temos evidência direta de que ela expressa um sentimento ou emoção de Michelângelo? Muitos achariam a posição estranha e tenderiam a rejeitá-la.

Uma razão mais forte para rejeitá-la teria que negar sua condição mais importante, argumentando que nem toda obra de arte envolve qualquer tipo de expressão sentimental ou emocional. Fazer isso envolveria mostrar algumas obras de arte assim. Podemos citar como exemplo algumas telas do pintor Piet Mondrian. Esse pintor compôs suas obras seguindo algumas rígidas regras autoimpostas, entre elas as de que as linhas das suas telas seriam apenas horizontais e verticais. Mondrian disse algumas vezes que não pretendia expressar emoção ou sentimento nenhum; ao contrário, afirmava que esse era exatamente um “problema” – isto é, algo a ser evitado – numa obra de arte. Ainda que se possa objetar que seus quadros expressam sentimentos como sobriedade e tranquilidade, seria difícil sustentar que essas qualidades não apenas eram emoções e sentimentos do próprio Mondrian, como também que é isso o que há de mais importante em suas telas. Sem contar que os relatos em primeira pessoa do próprio artista, que tanta força persuasiva deram à teoria expressiva da arte, nesse caso afirmarem o seu exato oposto.

Agora, todo artefato humano que expressa uma emoção ou sentimento é uma obra de arte? Também não. Uma pessoa que elabora uma lista de compras detalhada e funcional pode estar expressando um sentimento de ordem interior – ou algum outro sentimento parecido. Segue-se a consequência indesejável de que há ao menos um exemplo – talvez outros mais – de expressão de sentimento ou emoção por meio de um artefato humano que não é uma obra de arte. (Outros exemplos de artefatos que expressam emoções e não são obras de arte poderiam ser ventilados aqui).

Outro problema é o de que pode haver um desacordo radical entre as

emoções e os sentimentos que um artista porventura tenha visado a expressar e aqueles provocados no apreciador de sua obra. Um artista que visasse a expressar o seu sentimento patriótico pode acabar expressando muito mais do que tinha em vista. Como um exemplo politicamente explosivo desse tipo de desacordo, tomemos o filme *O Nascimento de uma Nação*. Concebido e dirigido por D. W. Griffith, parecia estar interessado em falar sobre os Estados Unidos da América durante e após a Guerra de Secessão e em expressar todo o apreço do seu realizador pelo seu país; no entanto, por apresentar personagens negros interpretados por pessoas brancas com os rostos pintados com a cor preta e por apresentar tais personagens como sexualmente agressivos e perigosos e com baixíssima capacidade intelectual, acabou expressando uma visão e sentimentos racistas, como desprezo e ódio com relação a pessoas negras; Griffith, ao tomar conhecimento de que milhares de pessoas reagiram violentamente contra a exibição do seu filme, sobretudo pessoas negras, produziu logo em seguida um filme chamado *Intolerância*, fazendo alusão à essa parte da recepção de *O Nascimento de uma Nação*. O diretor, portanto, parecia recusar a tese de que seu filme expressou sentimentos racistas, e que expressaria apenas sentimentos patrióticos, o que não parece ser o caso.

Por fim, assim como na afirmação central das teorias da imitação, a definição explícita da arte como a expressão do artista por meio de um produto humano pode incluir objetos que não são arte. Um discurso de um advogado no tribunal pode ter como objetivo fazer com que seus ouvintes fiquem indignados com o que acredita ser um genuíno caso de injustiça, expressando todo o seu próprio sentimento de indignação pessoal, mas dificilmente colocaríamos um discurso judicial sob o conceito de obra de arte. (Novamente, dizer que um discurso como esse é uma “obra de arte” parece estar apenas indicando algo como certa beleza ou capacidade de execução).

Todas essas razões indicam a necessidade de que a teoria seja revista. Para que alguma versão dela seja sustentável, seria preciso levar em conta todos os problemas mencionados e talvez outros que tenham passado pela sua cabeça. Pense um pouco sobre essas coisas. Há alguma outra dificuldade que a teoria expressiva da arte tenha que enfrentar? Existe algum modo de reformar a teoria para que ela supere tais problemas?

3.2.1 Collingwood e expressão imaginativa

Vimos anteriormente que alguns filósofos apresentaram a imitação como condição necessária para que determinado objeto fosse uma obra de arte. Diante da sugestão, poderíamos pensar: ‘há algum objeto que é comumente considerado artístico e que não envolve imitação?’. Em caso afirmativo, se temos uma tendência a valorizar o senso comum, teríamos uma razão para desconfiar da definição, pois muitas obras de arte são comumente classificadas como obras de arte e não envolvem imitação, como as pinturas abstratas, os poemas automáticos – poemas que envolvem sorteio de palavras – esculturas disformes etc. Contudo poderíamos prosseguir: ‘será que esse objeto comumente considerado arte é *apropriadamente* considerado arte?’. Com essas perguntas, começamos a tocar em questões sobre o modo como se faz filosofia e que tipo de condições esperamos que uma boa pergunta e/ou resposta satisfaça, afastando-nos um pouco mais da pergunta básica que nos fizemos – o que é arte?

Ao longo do século passado, boa parte dos filósofos começou a pensar que deveríamos olhar com maior cuidado para as nossas práticas linguísticas e culturais como um caminho de distinção sobre o que deve e o que não deve ser considerado uma boa definição. Não esperavam simplesmente catalogar e classificar as opiniões correntes, mas utilizá-las como um caminho de investigação e correção, pois parecia-lhes que o afastamento demasiado dessas práticas poderia ser sintoma de que se está sustentando uma posição falsa ou absurda.

Tomemos a teoria da arte de Collingwood como exemplo. No seu *The Principles of Art*, ele pretende responder muitas questões acerca da arte; entre elas, o que ela é. O filósofo sustentou que havia uma conexão bastante forte entre as obras de arte e processos de expressão das emoções humanas, motivo pela qual sua teoria ficou conhecida como uma teoria “expressionista” da arte.

Se quisermos oferecer uma formulação simplificada – talvez não excessivamente – da sua teoria, poderíamos dizer: para Collingwood, arte é expressão imaginativa. Mas essa é apenas a sua conclusão. O que mais nos interessa, na verdade, é investigar se as razões que ele oferece para sustentá-la são verdadeiras e se elas mantêm certa conexão apropriada com a conclusão.

A distinção mais básica que leva Collingwood a concluir que arte é

expressão imaginativa é aquela entre arte e ofício, que argumenta serem coisas completamente distintas. Ele argumenta que as teorias da arte até então tomavam a arte como pertencentes ao mesmo gênero dos ofícios. Um ofício seria uma relação entre meios e fins, uma atividade na qual uma matéria-prima é transformada de acordo com certos procedimentos e habilidades que podem ser aprendidas num produto previamente concebido. A sapataria, a marcenaria, a ourivesaria seriam exemplos de ofício.

Ele admite que pode haver casos de objetos que em determinadas circunstâncias e de determinado ponto de vista podem ser considerados arte. E também admite que todo artista precisa de certas habilidades muito próximas das dos ofícios (se não idênticas). Poetas compartilham com jornalistas a arte de utilizar palavras, assim como os pintores artísticos e os pintores de prédios o de utilizar tinta. Mas, como espera mostrar, o resultado desses usos é completamente diferente.

Não reconhecer essa diferença seria um erro para Collingwood. Ele esperava que ao analisar as práticas correntes da aplicação do conceito de arte na linguagem cotidiana, as pessoas viessem a reconhecer a diferença entre arte e ofício. Novamente, não significa apenas catalogar os usos, mas mostrar como há distinções, que tornam o funcionamento cotidiano da linguagem possível, que podem ser utilizadas como um guia da investigação. Com a análise concluída, esses usos sairiam mais precisos e renovados.

Após sua conclusão, esperava disponibilizar uma definição de arte “em sentido apropriado” e de ofício “em sentido apropriado”, a partir dos quais poderíamos avaliar nossas práticas linguísticas e sua adequação aos objetos.

Assim, sua primeira tarefa seria mostrar alguns casos de usos cotidianos “inapropriados” do conceito de arte. Ele os separa em dois tipos: a “arte de entretenimento” e a “arte mágica”. A arte de entretenimento seria uma arte que desperta uma emoção, que, por sua vez, não seria levada a nenhuma outra ocupação cotidiana, bastando-se a si mesma como algo valioso, que diverte e entretém. A maior parte do que é considerada arte, mesmo nos anos em que Collingwood publicou sua teoria, levava em conta tal independência. De acordo com o filósofo inglês, a experiência com tais objetos seria melhor descrita como “entretenimento”. Com isso, ele não pretende negar valor a esse tipo de objeto, mas simplesmente melhor distingui-lo de outros.

Um tipo diferente de uso inapropriado do conceito de arte ocorreria na

arte religiosa, um tipo específico do que chamou de “arte mágica”. Baseando-se em estudos antropológicos sobre a “arte mágica”, ele argumenta que, de um jeito diferente do entretenimento, a arte mágica desperta uma emoção que é conduzida para outras ocupações da vida.

No caso da arte religiosa, tal emoção pode conduzir a uma mudança de comportamento em relação às outras pessoas, por exemplo. Este tipo de arte estaria muito mais ligado ao modo de vida comunitário da qual emerge.

Nos dois casos, há um despertar de emoções, ou, para usar seu vocabulário, uma evocação de emoções. Na mágica, tal evocação teria uma relação estreita com a comunidade da qual emerge e agiria como um motivador da ação cotidiana. A emoção evocada pela arte patriótica pode levar à defesa do país, assim como a emoção evocada pela música religiosa é capaz de modificar costumes cotidianos, tais como o vestuário, a alimentação e a fala. No entretenimento, a evocação se encerraria na própria experiência, por serem irrealis e puramente imaginárias.

Nenhuma delas seria arte em sentido apropriado porque seriam melhor descritas como formas de ofício, se seguirmos sua proposta. Pois nos dois casos haveria a produção de emoções, por meio de procedimentos e habilidades passíveis de aprendizado, concebidos de antemão pelo seu produtor. A natureza da atividade do produtor de arte religiosa e de entretenimento seria a mesma da do produtor de sapatos e móveis de madeira.

Collingwood parecia estar bastante convencido da conexão necessária entre arte e emoção, embora a natureza dessa conexão seja um tanto controversa entre os leitores da sua obra. E evocação, de acordo com ele, pode ser comparada a um estímulo, mas não é idêntica. Adiante, a conexão é expressa nos termos de uma expressão – uma palavra que pode indicar uma natureza distinta da evocação. Ele passa então a caracterizar a natureza e o funcionamento da expressão, que seria o que distinguiria a arte dos ofícios. A distinção entre exprimir e descrever uma emoção é central nesse ponto da argumentação. Uma coisa é gritar gesticulando a frase “eu amo esse time!”, outra é dizer “tenho fome”. Na primeira, parece haver uma tentativa de comunicar um estado interior, ao passo que no segundo caso parece acontecer uma descrição de um estado interno. Mas o filósofo toma certo cuidado aqui, pois não está interessado nos modos costumeiros de comunicação de estados internos, como contorcer o rosto para exprimir dor. A essas reações, ele dá o

nome de revelações: modos não controlados de comunicação de estados interiores; a dor, por exemplo. Assim, a diferença entre propriamente exprimir e revelar um estado interior seria o controle e a consciência do controle dessa comunicação. De acordo com sua teoria, a capacidade de controlar e tomar consciência do controle das próprias emoções é a linguagem, que, num certo sentido, vai além do mero uso das palavras de determinado idioma.

O passo decisivo da argumentação é dado quando propõe que as definições de expressão de emoções, a arte em sentido próprio e a linguagem se referem à mesma coisa. A arte exprime emoção de uma maneira linguística, isto é, comunicativa. Aquele que exprime uma emoção está consciente de qual ou quais emoções quer exprimir, mas no caso da arte (em sentido próprio) o artista não a conhece de antemão; apenas no processo de expressão, isto é, enquanto produz a obra de arte, findando o processo apenas quando a obra for declarada pronta. Lembremos que se o artista conhecesse antecipadamente qual ou quais emoções sente e agisse conseqüentemente de modo a exprimi-la, estaria praticando o que Collingwood distinguiu como um ofício. Isso não quer dizer que o artista não tenha ciência de que irá produzir um romance ou um poema ou mesmo algum vislumbre do que deseja exprimir. O que conta é que uma emoção distinta e não conhecida de antemão se exprima por meio das técnicas disponíveis. Caso contrário, não será arte num sentido apropriado.

Ainda falta, para entender a definição, entendermos o que o autor entende por imaginação, afinal, é bom sempre ter em mente, arte é expressão imaginativa. 'Imaginativa' aqui significa basicamente que a arte em sentido próprio não diz respeito aos objetos públicos que associamos cotidianamente ao conceito – como um livro de romance ou poesia, uma tela na parede da sala ou no museu – mas às imagens mentais que se formam com ela, tanto do ponto de visto do seu produtor quanto do seu espectador, que exprimem emoções não conhecidas de antemão, por meio delas, enquanto são percebidas.

De sua definição de arte, Collingwood espera oferecer conjuntamente critérios para definição da arte e deduzir a partir dela critérios de avaliação. No seu modo de ver, “a definição de qualquer tipo de coisa é também a definição de uma boa coisa desse tipo”. Ou seja, um bom objeto é aquele que melhor se adequa à classe de objetos a que pertence. Um bom cão é tanto melhor quanto mais se adequa à classe dos cães. Além de contrariar um pouco do espírito declarado do projeto – seguir a linguagem cotidiana –, há claramente

uma confusão conceitual, notada por diversos críticos. Um cão pode satisfazer todas as qualidades que o definem como cão e, ainda assim, não ser um “bom exemplar” do animal – independente do que isso queira dizer. Não parece apropriado reunir numa mesma definição o que algo é e quais dos seus exemplares são bons. Alguém pode satisfazer as características que definem um jogador de futebol sem satisfazer as características que definem um bom jogador de futebol. Parece haver a necessidade de definições distintas a esse respeito, mesmo que possivelmente relacionadas.

Conforme mencionamos de passagem, Collingwood pretende utilizar a linguagem comum como um guia em sua investigação. Em vários momentos, no entanto, parece se afastar dela sobremaneira. Um exemplo sempre mencionado é sua caracterização das obras de Shakespeare como ofícios, mesmo que não pretenda diminuir sua importância com a distinção. Alguns veem aí um indício de problema, uma vez que Shakespeare parece ser um paradigma de boa obra de arte para muitos, coisa que as ideias do filósofo que estamos considerando nega.

Talvez o ponto mais importante a considerar em sua teoria seja sobre a conexão necessária entre emoção e arte, que parece supor que toda arte “em sentido próprio” envolve expressão. Que muitas obras envolvam algo como a expressão, muitos de nós estamos dispostos a concordar, mas que haja uma necessidade imprescindível de que ela ocorra, para que haja uma obra de arte, é um pouco mais questionável. Mais uma vez, poderíamos apontar para alguns objetos que são costumeiramente classificados como obras de arte e tentar argumentar que eles não parecem exprimir nenhuma emoção humana. Demos o exemplo do poema automático, mas poderíamos argumentar que mesmo alguns tipos de música ou textos em prosa, considerados obras de arte não, pretendiam exprimir nada e não parecem envolver nenhuma emoção aparente. Independente de se insistir na distinção da arte “em sentido apropriado” e “ofício” (o que tornaria a teoria mais afastada ainda da linguagem comum), haveria o problema de se mostrar que toda expressão é a expressão de uma emoção – e não, por exemplo, de uma ideia ou uma mensagem ou imagem etc. A certa altura de sua obra, o filósofo argumenta o que segue:

Em primeiro lugar, é fato que o discurso no qual é feita uma tentativa determinada de declarar

verdades mantêm um elemento de expressividade emocional. Nenhum escritor ou orador sério pronuncia um pensamento a menos que ele pense que vale a pena expressá-lo. O que faz valer a pena expressá-lo não é a sua verdade (o fato de que algo é verdadeiro nunca é motivo suficiente para dizê-lo), mas o fato de ser a única verdade que é importante na situação atual.

É difícil sustentar que todo discurso envolve uma consideração sobre o seu valor. Algumas frases são casuais e ditas irrefletidamente, poder-se-ia objetar. Além do mais, do fato de estarmos constantemente afetados por emoções não se segue que toda declaração que damos pretendem exprimi-la. Às vezes, parece se passar o exato oposto. Podemos fazer uma afirmação considerada importante, pensando que ela seja desimportante, por exemplo, quando damos uma informação sobre uma farmácia a alguém que necessita desesperadamente de um remédio, mas nos disse que gostaria apenas de se pesar. Quando descrevemos uma situação de fato no trabalho – como “há café na cafeteira” – podemos estar afetados por uma emoção que não mantém nenhuma relação com a declaração e porque temos o dever profissional de informar os demais; ou, ao contrário, é uma afirmação de pouca importância, que fazemos descuidadamente. Essa afirmação de que uma verdade só é dita a menos que seja muito importante na situação atual precisaria ser, no mínimo, refinada.

3.3. A teoria da arte como conceito aberto

Em meados do século passado, um importante filósofo que se dedicou a questões de Estética, entre elas às teorias da arte, pensava que havia um grande problema com *todas* as teorias da arte existentes. Desta vez, em vez de considerar a avaliação da afirmação central de uma família de teorias da arte, vejamos uma proposta específica e completa. No ensaio “O papel da teoria na estética”, Weitz (2004) tentou mostrar que todas as famílias teóricas anteriores, por exemplo, as teorias da imitação e as expressivas da arte, estava a investigar a definição do que seja arte, buscando por características importantes que sejam individualmente necessárias e conjuntamente suficientes. Ele argumenta o

seguinte:

A teoria estética – toda ela – está errada em princípio ao pensar que uma teoria é possível, uma vez que adultera radicalmente a lógica do conceito de arte. É falsa a sua contenda de que a ‘arte’ seja susceptível de uma definição real ou de outro tipo de definição verdadeira. A sua tentativa de descobrir as propriedades necessárias e suficientes da arte é logicamente ilegítima pela simples razão de que nunca aparecerá um tal conjunto de propriedades nem, conseqüentemente, a sua fórmula. A arte, tal como a lógica do conceito mostra, não tem nenhum conjunto de propriedades necessárias e suficientes; logo, uma teoria acerca dela é logicamente impossível e não apenas factualmente impossível. A teoria estética tenta definir o que não pode ser definido no sentido requerido.

Com essas afirmações Weitz não pretende dizer que todos os esforços anteriores foram inúteis. O que esses esforços teóricos não fizeram foi prestar atenção o suficiente na natureza distinta do conceito de arte, não obstante terem ajudado a iluminar porções de aspectos importantes da arte, de acordo com o filósofo. A afirmação final, segundo a qual a “teoria estética tenta definir o que não pode ser referido no *sentido requerido*” é crucial. O sentido requerido aqui é o de oferecer quais são as condições necessárias e suficientes que definem o conceito de arte, que daqui em diante passaremos a chamar de ‘tradicional’. Ele tentará mostrar que, por sua natureza distinta da grande parte dos conceitos, a definição de ‘arte’ precisa ser investigada e oferecida de outro modo.

Seguindo algumas ideias do filósofo Wittgenstein, Weitz tentará mostrar que o conceito de ‘arte’ é *aberto*. Um conceito seria aberto quando os objetos que ele define podem ser reconsiderados ao longo do tempo. Num determinado momento do tempo, digamos, o século XVIII, o conceito de arte poderia se referir a certos objetos; no século XIX, poderia já ter incluído outros; e assim sucessivamente. Da mesma maneira como Wittgenstein tentou mostrar que o

conceito de “jogo” é aberto, isto é, que não há características comuns a todos os jogos que possamos descrever, ao oferecer uma definição tradicional, Weitz pensou que o conceito de arte era do mesmo tipo. Ao tentar oferecer uma definição tradicional, os filósofos teriam, na melhor das hipóteses, capturado um estado momentâneo do conceito de arte.

Veja que as teorias da imitação não providenciam muito bem um lugar para *A Fonte*, de Duchamp. O mictório utilizado para compor a obra não imita a aparência de nenhum outro objeto. Ele é um objeto idêntico aos outros. Sendo assim, não se ajusta bem à teoria da imitação, obrigando-nos a escolher entre considerá-la um exemplar de obra de arte ou se a definição não procede. O mesmo vale para a teoria expressiva. Assim, pairaria uma sensação de que todas essas definições são arbitrárias e nada mais que isso. Mas se Weitz está correto, isso não deveria nos espantar, pois seria da própria natureza do conceito de arte a mudança ao longo do tempo. Se mantivermos o conceito “aberto”, então é de se esperar que possamos acomodar todas essas obras apropriadamente sob o conceito de arte. Mas se obras tão diferentes entre si, como *Davi*, *Guernica* e *A Fonte* não possuem nenhuma característica ou conjunto de características comuns que sejam suficientes para distingui-las dos outros objetos, o que justificaria dizer que todas elas são exemplares de “obras de arte”?

Weitz responde que são suas *semelhanças de família* que nos permitem justificar a aplicação do conceito – uma ideia que ele também retira do pensamento de Wittgenstein. Quando estamos diante de uma família, temos a nítida impressão que todos os indivíduos que a compõem possuem certas características comuns. O formato dos olhos, do nariz, a altura, o modo de caminhar e falar, a cor dos cabelos e da pele, o tamanho das mãos etc. etc. Digamos que o nome dessa família seja Oliveira e suponha também que eles não tenham nenhuma característica partilhada por todos os seus membros. Seria estranho que suspendêssemos nossa “classificação” dos membros dessa família como pertencentes à família Oliveira apenas porque eles não têm uma característica partilhada por todos. Ao contrário, além de receberem o nome da família de seus pais, os Oliveira possuem um conjunto de características que os tornam muito parecidos entre si. O seu João Oliveira tem olhos azuis e mãos grandes; sua filha, Joana Oliveira, tem olhos azuis também, mas tem mãos de tamanho médio, ao contrário do seu irmão, Thiago Oliveira, que não possui olhos

azuis, mas possui mãos grandes; e assim por diante.

Essas *várias* características compartilhadas entre primos, tios, avós etc., formam o que se costuma chamar de semelhanças de família. Weitz pensa que as obras de arte se relacionam de modo parecido entre si. Elas não possuem nenhuma característica comum, mas um conjunto de características compartilhadas umas com as outras. Para simplificar, pense em cinco obras de arte, que chamaremos de 'A', 'B', 'C', 'D' e 'E'. Imagine que a obra 'A' compartilhe alguma semelhança com a obra 'B' e com a obra 'E'. Imagine também que a obra 'E' compartilhe semelhanças além de 'A', com 'C' e 'B'; que 'C' e 'B' não compartilham nenhuma semelhança entre si. 'C' compartilha semelhanças com 'E' e 'A' somente, enquanto 'B' compartilha semelhanças com 'A' e 'E'. Essas semelhanças justificariam considerá-las parte de uma família, permitindo dar a elas um nome comum – obras de arte.

Dessa maneira, poderíamos evitar “fechar” o conceito de arte e impedir que ele exclua novos objetos, como *A Fonte*. Para que ela e qualquer outro objeto possam ser incluídos no conceito de arte, bastaria encontrar uma característica relevante que as assemelhe a outras obras de arte, destacada por teorias como a expressiva e da imitação, para então inclui-la na família. Note que não há qualquer necessidade de busca por condições necessárias e suficientes aqui; apenas que faça parte da família por meio de algumas características compartilhadas.

Weitz levanta ainda outro ponto a favor de sua teoria. Ele diz que as definições tradicionais prejudicam a criatividade necessária à arte, ao passo que a sua teoria poderia ter um papel estimulante para a produção de novos tipos de obra de arte. Em vez de ter um papel limitador, como o filósofo pensa que essas teorias tiveram, as novas teorias poderiam contribuir positivamente para a produção artística, recebendo bem a inovação. A teoria de Weitz é bastante sedutora, em primeiro lugar, porque ela oferece uma explicação para a nossa aparente incapacidade de oferecer uma definição tradicional bem-sucedida do conceito de arte. Não diz que foi somente uma falha generalizada dos filósofos anteriores, diz apenas que eles não prestaram atenção a um ponto crucial da natureza desse conceito – que ademais contribuíram conjuntamente para esclarecer. Além disso, sua proposta providencia uma maneira de resolver vários desacordos sobre o que deve ou não contar como arte.

Para avaliar essa teoria precisamos considerar ao menos duas coisas.

Primeiro, arte é realmente um conceito aberto? Segundo, uma definição tradicional – que “fecha” o conceito de arte – é realmente prejudicial à produção de novos tipos de arte? Alguns filósofos pensaram que nenhuma das coisas realmente é o caso. Sustentaram que há um passo a ser observado com mais atenção: do fato de todos os filósofos não terem encontrado uma definição tradicional, não se segue que a definição tradicional não possa ser encontrada. Pode ser o caso que todos estejamos a falhar até o momento, o que não seria nenhuma novidade nas tentativas humanas de conhecer melhor o mundo. Mas Weitz tenta ir além, argumenta que todas as tentativas falharam e devem necessariamente falhar, pois violam a natureza própria do conceito de arte, produzindo engano.

Alguém convencido da teoria de Weitz poderia nos convidar a observar os vários usos aparentemente bastante razoáveis do conceito de arte e indicar que aí está a evidência de que esse conceito é aberto. A dificuldade é que alguém poderia objetar que é exatamente aqui que uma teoria tradicional poderia nos ajudar, pois ajudaria a separar quais atribuições do conceito de arte são ou não apropriadas, pois algumas dessas aplicações podem ser indevidas.

Uma objeção poderia minar um pouco a confiança nas semelhanças de família de nos ajudarem a definir conceitos que sejam por natureza abertos. Em algum ponto, todos os objetos podem ter alguma semelhança entre si. Um exemplo clássico, oferecido por Warburton (2002, p. 81), mostra como muitas semelhanças de família podem ser simplesmente irrelevantes: o *Empire State* (um prédio localizado em Nova York) e uma agulha de costura são semelhantes, pois são ambos pontiagudos e feitos de material inorgânico (poderíamos acrescentar que são produções humanas, envolvem uso de metais etc.). Então, seria necessário acrescentar algum outro critério que delimitasse quais são as semelhanças relevantes que permitem incluir um objeto apropriadamente no conceito aberto de arte. As várias teorias, críticas, testemunhos e assim por diante poderiam ser uma ótima forma de encontrar quais seriam essas características relevantes, na medida em que se propõem a iluminar aspectos importantes das obras de arte, não fossem os numerosos desacordos entre si. Uma vez que não são consistentes, isto é, algumas dessas afirmações são conflitantes, quais critérios de semelhança considerar e quais desconsiderar? É uma questão espinhosa.

Alguns críticos da analogia com as semelhanças de família a

consideram muito limitada e escolhem essa via para avaliar a teoria de Weitz. Um conjunto de semelhanças cruzadas pode ajudar a reconhecer uma família biológica, mas nem sempre esse é o caso. Muitas famílias biológicas diferem tanto entre si em características fenotípicas, em virtude da variação genética, que fica difícil o reconhecimento delas apenas por características visíveis. Então é falso que um conjunto de semelhanças sempre permite reconhecer as famílias biológicas. É possível dizer ainda que há outra maneira de definir uma família sem recurso a características visíveis: por relações genéticas ou de origem. Podemos saber se uma pessoa é filha de outra recorrendo a um teste de semelhança genética, como podemos simplesmente traçar uma árvore genealógica; tais coisas mostram que há maneiras mais precisas de delimitar uma família.

Mais uma vez, há problemas a se considerar. Muitos não estão dispostos a abandonar a tarefa de buscar uma definição de arte, que nos permita distinguir adequadamente esses objetos, pois não acreditam que as falhas do passado tornam inevitável falhas no futuro, ou, pelo menos, acreditam numa maior aproximação ao acerto. Muitos filósofos discutem esses pontos desde então e, como de costume, tomaram essas considerações para reformar ou avançar novas propostas de teorias da arte.

4. Teorias da beleza

A noção de beleza, como você deve se lembrar da introdução, é bastante corriqueira. Se você prestou bem atenção na introdução desse texto, deve ter notado que as duas perguntas que aparecem conduzem a caminhos teóricos diferentes. Se nos perguntamos o que faz uma obra de arte, digamos, *Garota de Ipanema*⁹⁴, ser uma bela canção; ou se não sentimos nada de especial ouvindo tal canção, e pensamos que outra canção de gênero ou estilo muito diferente nos faz *sentir* a beleza, então daremos respostas diferentes à questão 'o que é a beleza?'. No primeiro caso, parecemos estar buscando algo que pertence à própria canção, isto é, a beleza estaria objetivamente nela. Se nos perguntamos o que ela nos faz *sentir*, então a canção pode no máximo desencadear alguma experiência *interna* e *particular* em quem a ouve. Com

⁹⁴ Canção composta por Tom Jobim e letrada por Vinícius de Moraes em 1962.

poucas adaptações, podemos estender o mesmo exemplo às demais artes e, quem sabe, às paisagens e objetos naturais.

Foram poucos os filósofos que tentaram oferecer uma resposta para a pergunta ‘o que é a beleza?’ em termos objetivos, como se a beleza fosse uma propriedade totalmente independente da nossa experiência pessoal. Os filósofos que seguiram nessa direção tiveram como inspiração intelectual alguns marcos teóricos apresentados pelos diálogos de Platão.

4.1. Platão

Platão é bastante conhecido como o primeiro filósofo a elaborar uma definição de arte, como algo distinto de atividades técnicas ou ofícios, tais como a sapataria e a marcenaria. E também como um filósofo cuja opinião era extremamente negativa ao que atualmente chamamos de belas-artes. Como tudo o que diz respeito a ele, no entanto, as coisas nunca são tão simples. Assim, adotaremos a visão mais tradicional acerca das opiniões do filósofo sobre a arte, recomendando, ao mesmo tempo, que ela seja vista com certa cautela.

Essa visão se deve principalmente a certa interpretação largamente aceita daquela que é conhecida como sua principal obra, o diálogo chamado *A República*. Sócrates, a personagem protagonista da maioria dos seus diálogos, defende nela uma posição drástica em relação aos poetas. Numa cidade justa, afirma Sócrates, os poetas deveriam ser nada menos que banidos, isto é, deveriam ser expulsos e impedidos de retornar.

Que razões Sócrates oferece para defender uma posição tão drástica?

A primeira é a de que os poetas, tendo Homero como o representante máximo deles, não seriam bons exemplos morais para crianças e jovens, já que frequentemente os cantos dos poetas apresentam deuses e homens agindo de maneira imoral. Dessa forma, ouvir tais cantos teria uma ação maléfica sobre a alma dos jovens em formação; o que, por sua vez, uma cidade justa não deveria permitir.

Mais adiante, Sócrates apresenta a célebre Teoria das Formas, que mais tarde será a base para uma rejeição ainda mais fundamental da arte. De acordo com esta teoria, grosso modo, todos os objetos com quais temos contato diariamente não seriam propriamente “reais”, mas meras sombras ou vagas lembranças de outros, estes sim, reais: as Formas. Ao modo como os objetos

com os quais temos contato sensível se relacionam com as ideias, Sócrates deu o nome de “participação”. Assim, se vemos uma cadeira diante de nós, de acordo com esta teoria, não estaríamos diante de uma cadeira real, mas de uma sombra ou vaga lembrança da cadeira real: a “a Forma da Cadeiridade”.

Se Sócrates tivesse em vista somente o argumento moral contra a permissão para que os jovens ouvissem poesia, poderíamos pensar que basta trocar o “conteúdo” desses cantos, proibindo apenas aqueles que apresentam imoralidades. Seria, se não houvesse um motivo ainda mais fundamental para banir não apenas a audição de poesia, mas os próprios poetas da Cidade Justa. No último livro de *A República*, Sócrates oferece razões que não pertencem apenas ao “conteúdo” da poesia – e, por extensão, de todas as belas-artes –, mas à sua própria natureza. Para expor seu ponto de vista, com seus diversos detalhes sempre intrigantes, ele compara as obras de três diferentes tipos de produtores, a saber, a do demiurgo divino, as do marceneiro e as do pintor, tendo como demonstrada a Teoria das Formas.

Começemos com o marceneiro. Os marceneiros, assim como todos os artesãos, quando estão no processo de produção de uma cadeira, por exemplo, têm sempre em vista a “Forma da Cadeiridade”. Essa Forma serve como modelo e orienta todos os procedimentos do artesão para que possa chegar ao seu resultado: a cadeira. No entanto eles não podem produzir uma cadeira real, uma vez que esta não poderia ser produzida por procedimentos humanos. As Formas, os únicos objetos reais, afirma Sócrates, só podem ser produto de algum deus. E quanto ao pintor, o que ele faz? O pintor, quando pinta uma cama ou uma árvore – isto é, representa uma cama ou uma árvore por meio da pintura – não tem em vista a própria Forma da cama ou da árvore, mas busca representar um objeto que, por sua vez, apenas possui participação nas Formas.

O pintor realiza o que Sócrates chamou de “imitação” (mímese). O resultado do seu trabalho não é um objeto real (bem como não era o do artesão), mas a produção de uma aparência dos objetos – já eles mesmos aparências das Formas. Enquanto a cadeira seria uma cópia imperfeita da Forma da cadeira, a representação de uma cadeira seria a cópia de uma cópia. A conclusão a que chega é a de que a atividade do artesão está duas vezes afastada, ao passo que a do pintor está três vezes afastada da realidade. (Os povos gregos não possuíam o algarismo zero no seu sistema numérico; com efeito, iniciavam a contagem uma unidade além da que estamos habituados).

Mais uma vez, Sócrates relaciona a natureza da arte (a imitação de aparências) a aspectos morais considerados de extrema importância. A natureza da arte prejudicaria a formação da alma na Cidade Justa – sobretudo do que chamou dos seus “guardiões” – não apenas por seu afastamento da realidade – o que seria por si indesejável – mas pelos efeitos da essência dos objetos artísticos sobre o caráter. Uma das razões é que essas obras são mutantes. No caso da poesia isso era bastante visível, pois a poesia assumia feições diferentes de acordo com o rapsodo encarregado da sua execução e da ocasião. A prática da recitação da poesia incentivaria a absorção do caráter dos personagens narrados e do estilo de imitação. O resultado poderia ser o de que o jovem guardião tivesse absorvido qualidades indesejáveis – “infantis” ou “femininas” –, além de ter uma personalidade sumamente instável e indefinida – algo indesejável para a Cidade Justa que Sócrates e seus interlocutores estão construindo como uma maneira de investigar qual é a essência da justiça.

Até aqui parece que opinião difusa segundo a qual a visão de Platão era bastante negativa em relação à arte, nos exemplos da poesia e da pintura, se confirma. Mesmo que identifiquemos Sócrates como o porta-voz das posições filosóficas pessoais de Platão, no entanto, como muitos bons intérpretes de sua obra sugerem, não resulta claro qual é a sua posição. Não apenas por serem abundantes as interpretações plausíveis de *A República*, mas entre outras coisas, por serem as próprias obras do filósofo belos exemplares de imitações – como outros filósofos da mesma época, como Aristóteles, sublinharam.

Atualmente muitos tomam a teoria apresentada por Sócrates em *A República* como a primeira teoria da arte da tradição filosófica, querendo dizer com isso que essa foi a primeira resposta para a pergunta ‘o que é arte?’. Como vimos, arte é imitação das aparências. Mas seria muito difícil sustentar hoje que essa resposta seja suficiente. Imitar as aparências – independente das nuances e do sentido preciso do que se entende por “imitar” – e mesmo que estejamos de acordo com a teoria das ideias, seria uma condição necessária, mas dificilmente suficiente para definir arte, como já vimos. De fato, parece que seria possível imaginar imitações, no sentido que Sócrates as define, que não sejam arte – como um mapa, por exemplo. Algumas obras não figurativas que são atualmente consideradas arte, pelo menos a princípio – isto é, que não procuram trabalhar a partir da aparência cotidiana das paisagens, dos espaços, dos objetos, das pessoas etc. – não seriam arte de acordo com a definição. Pinturas célebres

como as de Jackson Pollock não seriam consideradas arte, já que não envolvem imitação, ao menos não no sentido mais tradicional.

É importante tomar alguns cuidados aqui. Nem Platão nem qualquer outro grego podem ter oferecido uma definição sobre o que é a ‘arte’, como atualmente costumamos nos perguntar. Esse conceito, que abrangeria objetos tão distintos quanto esculturas, pinturas, poesias e músicas, pode estar sendo levado em consideração naquelas obras de algum modo, mas pode ser o caso que o que se tinha em vista fosse a definição de atividades particulares. Ou, ainda, que o conceito culturalmente mais difuso que reuniria todas essas atividades não fosse o de “arte”. Se isso estiver certo, não haveria uma tentativa de definir todas as “artes”, mas exercícios de definição e avaliação de atividades específicas, como a produção e audição de poesia. Mais uma questão para ficar de olho e aprofundar nos seus estudos.

A beleza é uma noção que variou muito ao longo da história e da história da filosofia, e Platão também é pioneiro aqui, por ter sido um dos primeiros a tematizar o que a beleza é. Embora o que chamamos de obras de arte – como uma bela canção a tocar nos seus fones de ouvido – não tenha aparecido no texto mais célebre em que trata do assunto, *O Banquete*. O principal assunto a ser discutido pelas personagens nessa obra é o amor. No momento mais importante do diálogo, cada uma das personagens faz um discurso acerca dele. Uma mulher, vidente e sacerdotisa, chamada Diotima de Mantinea, faz um discurso muito famoso no qual narra um caminho de “subida” até a Forma da Beleza. O tratamento que ela dá a beleza, como pode já ter notado, é similar ao das demais Formas, constituindo uma aplicação da mesma teoria. Nesse caminho de subida, nos é dito que todas as características sensíveis pelos quais a beleza é percebida – num belo corpo humano ou nas ciências, por exemplo – vão sendo deixadas para trás, até que a pessoa no processo de “ascensão” contemple apenas a própria Forma da Beleza.

Alguns autores sugerem que, em outros diálogos, Platão busca investigar o que todas as coisas belas sensíveis têm em comum, não estando interessado somente na Forma da Beleza, portanto, mas também nas características que as entidades devem satisfazer para participar da Forma, embora esse seja um ponto controverso. Aliás, como já dizíamos, as coisas em Platão são extraordinariamente difíceis. O que pode ser tomado como um desafio e um convite estudar suas obras e pensar criticamente sobre o que elas

propõem.

A maneira platônica de ver a beleza foi muito comum até mais ou menos o século XVIII, quando houve uma mudança geral de orientação sobre a questão.⁹⁵ Uma boa razão para essa reorientação filosófica é a de que parece ser muito difícil um acordo unânime sobre a beleza; outra é a de que parece inexistir um conjunto verificável de condições que garantam que determinado objeto ou ação serão sempre belas.

Mesmo não sendo tomada como exclusivamente pessoal pela maior parte dos filósofos, a beleza passa ser vista primariamente como uma experiência, isto é, como um processo pessoal que ocorre a uma pessoa quando ela percebe um objeto. Assim, começa-se a pensar que é na análise dessas experiências associadas à arte e ao belo que se encontra a chave para as respostas. Daí em diante, a noção de *gosto* assume uma posição central e é acerca dela que falaremos brevemente agora, passando em revista e buscando avaliar o pensamento de dois expoentes dessa nova orientação filosófica.

4.1. Hume

David Hume, um eminente filósofo escocês nascido em 1711, publicou, em 1757, uma importante obra, cujo título é “Sobre o Padrão do Gosto”. Nela, Hume acredita ter algumas boas explicações para o problema de saber por que percebemos alguns objetos como belos. Ele espera delimitar um aspecto importante do que chamou de “natureza humana”; aparentemente, Hume estava interessado apenas em descrever quais fatores levam todos os seres humanos à experiência da beleza.

Neste ensaio, Hume reconhece de saída que há muito desacordo

⁹⁵Em seus futuros estudos sobre estética, vale a pena dedicar alguma atenção a essa mudança de orientação na questão, que afetou não apenas a estética, mas todas as áreas da filosofia. É nesse período que Baumgarten propõe a nova ciência da estética, que já vimos o que significa para ele. O desenvolvimento posterior da disciplina, curiosamente, deve muito pouco ao sentido preciso que Baumgarten procurou lhe dar. Houve um declínio cada vez maior das formulações que viam na “beleza” e na “arte” algo objetivo, isto é, pertencente aos próprios objetos. Outras noções como as de “sublime”, “grotesco”, “pitoresco” passaram a receber tanta atenção quanto a noção de beleza.

genuíno entre os seres humanos sobre o que experimentam ou não como sendo belo – ou, em poucas palavras, há muito desacordo de gosto; mas esses desacordos seriam resultado de circunstâncias acidentais nas quais os indivíduos se encontram. Ele sabia que há muita gente que duvida da possibilidade de “discutir o gosto” de alguém; contra tal ceticismo, argumenta que seria absurdo, pois isso implicaria ter de aceitar que não podemos distinguir obras que são melhores do que outras. Pense, por exemplo, em músicas que são vistas pela maioria das pessoas como medianas e outras que são consideradas por elas como belíssimas. Se fôssemos incapazes de distinguir uma música mediana de uma excelente, então tal ceticismo estaria justificado; como ele acredita fortemente que não somos, conclui que é um absurdo sustentar tal ceticismo. Para mostrar por que está certo, Hume apresenta obras reconhecidamente desiguais em mérito e supõe que até um cético aceitaria que uma obra é superior à outra.

Hume não pensa por isso que estamos aplicando critérios independentes das nossas experiências mais comuns quando reconhecemos certas obras de arte como superiores a outras. Haveria um padrão passível de descrição detalhada do que todos os “povos” e “épocas” consideraram belo. Esse padrão estaria na base de todas as experiências da beleza. Agora bastaria considerar tudo o que as pessoas, em todos os povos e épocas, experimentaram como belo e encontrar o padrão, certo? Nem tanto. Hume nega que todas as experiências que as pessoas declaram como agradáveis contam como indícios a serem levados em conta na generalização que nos daria acesso ao padrão. O filósofo escocês pensava que apenas as experiências acontecidas em determinadas situações e com as disposições emocionais apropriadas, tais como a serenidade, o recolhimento e a atenção, podem contar como parte do inquérito sobre o padrão do gosto, isto é, as condições que produzem a experiência do belo. Assim, experiências agradáveis de uma obra de arte que sejam oriundas da inveja, de modas passageiras e pressões de grupo não contariam. Ademais, Hume julga constatar que há indivíduos mais capazes de reconhecer uma boa obra do que outros; usa uma analogia para apoiar seu ponto de vista, afirmando que, assim como a maioria das pessoas, os degustadores de vinho possuem a qualidade do paladar, mas é notório que os degustadores são muito mais capazes de distinguir os vinhos bons dos ruins.

Embora seu foco recaia sobre a experiência, Hume argumenta ainda

que algumas características dos próprios objetos desencadeiam esses processos internos, muito embora não especifique quais são. Aparentemente, então, há um encontro entre as características estáveis da natureza humana e certas propriedades dos objetos na produção da experiência apropriada do gosto. Curiosamente, no entanto, Hume encerra seu célebre ensaio dizendo que outros dois fatores podem contribuir para experiências apropriadas de gosto, além da disposição e da situação: o “temperamento” e a “idade” do indivíduo.

Para avaliar os argumentos de Hume, teremos que considerar e interrogar vários pontos. Os principais parecem ser: a) se a beleza é uma questão exclusivamente empírica, isto é, poderemos resolvê-la simplesmente observando o comportamento humano?; b) serão lícitas as exigências que propõe a uma experiência “apropriada” do gosto? ou esta é apenas uma forma mais ou menos arbitrária de encontrar um padrão?; c) do fato de alguns indivíduos serem capazes de distinguir melhor certas propriedades sensíveis dos objetos, como o sabor do vinho, se segue que são mais capazes de avaliar que um objeto, como o vinho, é *melhor* do que outro?; d) qual é mais precisamente o papel dos próprios objetos nessa experiência? todos eles podem desencadear a experiência da beleza?; e) qual é o papel da idade e do temperamento nessas experiências apropriadas de gosto? em que medida pode haver variação no padrão de gosto e ainda assim termos um padrão?; f) qual é o papel do treinamento ou da educação para que um indivíduo tenha uma experiência de gosto apropriada?; h) Essa experiência apropriada se refere apenas às obras de arte ou também se estende a outros objetos humanos, bem como à natureza?

Essas, entre outras questões, foram discutidas por muitos filósofos desde então. Como se pode ver, nossas questões cotidianas deram e dão origem às mais complexas buscas por respostas e exigem resposta daquele que quer tentar encontrar um caminho para resolver à pergunta inicial: o que é arte? O que é o belo? Ou, como o próximo filósofo a considerar, elas obrigam a repensar os pressupostos da própria pergunta para, então, buscar novas respostas?

4.2. Kant

Dentre todos os filósofos citados até agora, aquele que talvez exija a maior dedicação para o entendimento do seu pensamento sobre as questões

que nos interessam aqui é Immanuel Kant. Nasceu em 1724 e viveu na cidade Königsberg até falecer, em 1804. Se não tivermos uma boa ideia do seu projeto filosófico geral – e compreendê-la já é uma atividade desafiadora – ficaremos um tanto atônitos com suas propostas. O que poderemos oferecer aqui são algumas considerações preliminares e esquemáticas, mas que esperamos não se afastarem demasiado do seu pensamento.

Diferente de Hume, Kant não pensava que a experiência do belo – a qual chama juízo estético de gosto – é uma experiência que podemos resolver por vias exclusivamente empíricas. Não bastaria observar o comportamento das pessoas; seria preciso analisar as condições de possibilidade dessa experiência, presentes em todo sujeito dessas experiências. A experiência do gosto foi investigada por Kant num livro chamado “Crítica da Faculdade do Juízo”, que faz parte de uma trilogia, intitulada “projeto crítico”. Quando Kant fala em “faculdade”, ele espera designar uma capacidade ou uma atividade da mente humana, embora em sua época não parecesse despropositado falar numa “faculdade de pular, de jogar etc.”.

Quando fala em “juízo”, Kant indica uma atividade intelectual que busca combinar representações singulares e universais. Assim, no juízo “esta flor é uma rosa”, temos a indicação de um termo singular (esta flor) que deve ser particularizado mediante uma associação com um termo universal (rosa). A “faculdade de julgar” seria a capacidade humana de fazer juízos, relacionando algo singular com um conceito. Já o “ajuizamento” seria, grosso modo, o processo de formação do juízo, processo de bastante importância no juízo de gosto.

Há dois tipos de juízo que nos interessam. O primeiro o que chamou de “juízo determinante”, que se encaixa no exemplo citado antes. Mas há também outro tipo de juízo, que é o “juízo reflexionante”; ele ocorre quando temos um objeto diante de nós para o qual não consideramos possuir um conceito apropriado para associá-lo e, assim, particularizá-lo. No juízo reflexionante, todos nós pensamos, argumenta Kant, que há algum conceito que deve “se encaixar” no objeto que temos em vista, mas não sabemos ainda qual é. Note que há uma expectativa envolvida no juízo e nada nele garante que de fato vamos encontrar o conceito. Desse modo, a maior parte dos nossos juízos cotidianos seriam do tipo determinante, mas às vezes nos ocorrem os reflexionantes.

O que a experiência do belo teria a ver com os juízos? Os juízos se

manifestam de modo verbal, numa sentença qualquer. “Esta flor é uma rosa” é uma manifestação do juízo de um sujeito que tem uma flor particular em mente. Mas na experiência do belo se passaria da mesma forma? Sim, para Kant a experiência do belo – um juízo reflexionante específico, o de gosto – não é uma experiência puramente física; ela precisa ser articulada também em juízos. Outro ponto importante é que grande parte do seu prazer vem do seu compartilhamento com outros indivíduos, que não podem sentir os estados físicos das outras pessoas.

O juízo de gosto é bastante intrigante, de acordo com Kant, porque ele não se articula genuinamente em juízos como “esta rosa é bela”, pois esse seria um juízo determinante, que não teria relação com a experiência genuína do belo. Nos juízos de gosto, o sujeito não encontra um conceito apropriado para o objeto que temos em vista e reside aí grande parte do prazer que eles causam. Também não há nenhuma regra geral que indique quando esse tipo de juízo ocorrerá diante de um objeto. Dois objetos podem ser produzidos de acordo com os mesmos princípios, serem inclusive muito similares, e ainda assim um deles pode proporcionar uma experiência do belo e outro não. E aí está contida, de acordo com o filósofo alemão, outra fonte do prazer causado pelo juízo de gosto – o fato de não sermos capazes, enquanto ocorre a experiência, de desvendarmos sua regra de composição.

Por exemplo, quando você se dá conta da maneira mais ou menos repetitiva como filmes de Hollywood são organizados e produzidos, de acordo com esta forma de pensar, uma condição necessária para uma genuína experiência do belo se perde. Isso não significaria que eles não sejam divertidos, agradáveis e mesmo fonte de grande prazer pessoal, mas não seriam capazes de oferecer uma experiência do belo. Outro exemplo poderia ser a de um prédio público da sua cidade. Muitos turistas que ainda não o conhecem e não são capazes de “enxergar” seus procedimentos enquanto o observam e caminham por ele podem sentir uma experiência do belo, mas tão logo se tornem familiares e seus procedimentos de produção sejam conhecidos, tais obras deixam de proporcionar a mesma experiência. As obras e paisagens capazes desencadear a experiência do belo são aquelas nos aparecem como “espontâneas” e “desprovidas de esforço”, onde sentimos que certa irregularidade em sua composição impede que elas sejam rapidamente classificadas e sejam, consequentemente, fonte de tédio.

Mesmo sendo essas experiências que ocorrem a cada um dos sujeitos, Kant argumenta que elas não são propriamente individuais. Ele não está dizendo que, por exemplo, um romance literário não possa ter um grande valor e nos provoque grande prazer, enquanto outras pessoas não veem a menor graça nele e nenhuma qualidade nele. Mas aqui estaríamos falando de outras experiências, como o puro prazer, a nostalgia, a familiaridade etc. Dizer sobre uma pintura “isto é belo para mim” seria algo despropositado, além de soar “estranho e risível” para ele.

Em tempos de tamanha oferta de objetos “culturais” é difícil não estranhar a proposta kantiana. Mas pense numa paisagem linda, como as de Fernando de Noronha, na região nordeste do Brasil ou em edifícios como o Kremelin, na capital da Rússia. Ficaria um pouco mais difícil dizer que elas são belas “para mim”. Pense no contrário, é fácil imaginar alguém dizendo sinceramente “esta praia em Fernando de Noronha é horrível!”? Não parece fácil para aquele que escreve estas linhas, embora não seja de se duvidar que possa ocorrer por aí.

Agora, para entendê-lo um pouco mais, precisamos penetrar numa análise um pouco mais árida, na organização interna do juízo estético de gosto. Kant nos diz que há duas considerações a serem feitas acerca da “qualidade” do juízo. Se eles são “estéticos” ou “lógicos”. Os juízos lógicos são aqueles que têm a pretensão de representar algo objetivo acerca do mundo. Um exemplo de juízo lógico seria articulável na frase “o computador é preto”, onde parece haver implícita a pretensão de dizer que essa representação coincide com o mundo externo, isto é, independente dos nossos processos mentais individuais. Os juízos estéticos visam representar a maneira como somos afetados pelo mundo externo; por meio deles, descrevemos como somos afetados, sobretudo por essas próprias representações. Um exemplo: quando digo “que café amargo!”. Nessa frase, estaria articulada uma representação de um objeto externo (o café) e uma propriedade universal (o amargo), além de como essa representação me afeta.

Os juízos estéticos de gosto (um subtipo de juízo estético, portanto) são aqueles que dizem respeito ao sentimento do belo. Por não ser propriamente individual, este juízo sempre tem a pretensão de representar algo sobre o mundo, embora sua correspondência com ele não seja garantida. Quando alguém expressa num juízo estético de gosto, digamos (supondo que está

ocorrendo uma experiência do tipo que tem em vista), “esta escultura é fantasticamente tática!” há uma liberação de prazer desinteressada. Isso quer dizer que, quando esse tipo de experiência está ocorrendo, nenhuma referência externa importa, nem a do objeto que porventura esteja sendo representado, nem a da moralidade da representação ou do sentimento que causa; tal experiência não nos torna mais felizes nem aumenta nossa indignação frente aos problemas do mundo. Se isto estiver correto, para que tenhamos uma experiência genuína do belo diante de uma suntuosa catedral colonial, deveríamos apagar de nossas mentes o fato do horror à escravidão; da mesma forma, diante de um quadro ou uma estátua que represente uma figura histórica que não aprovamos, precisamos esquecer por alguns instantes ao que se referem, para que ela ocorra. Podemos considerá-las depois, seguindo o raciocínio kantiano, mas não *durante* o juízo estético de gosto.

Apesar de o juízo de gosto ser individual, sua pretensão seria universal. Negando o empirismo de Hume, Kant argumenta que tal juízo não é um apenas um sentimento, embora os envolva. E esse é o ponto central de sua visão sobre a beleza. Não é nenhuma relação simples de um objeto com os nossos sentidos que produz o juízo de gosto. Esse não é um juízo estético qualquer, como aquele oriundo do encontro do café sem açúcar com o nosso paladar, que nos faz articular o juízo “que café amargo!”. Faz parte dos juízos de gosto esperar que todos os demais concordem comigo; se experiencio uma casa como bela, há algo nessa experiência, de acordo com Kant, que nos leva a esperar uma concordância universal.

Essa sutileza advém da compreensão de um ponto crucial desse tipo experiência. Nela ocorre o que ele chamou de “reflexão”, um modo especial de identificar e comunicar esses estados especiais. Nela, as nossas “faculdades” funcionam de modo distinto do habitual e nenhuma delas prevalece sobre a outra. O “entendimento”, a faculdade envolvida nos juízos determinantes, e a imaginação, uma faculdade envolvida nas “sínteses” (nas “junções” entre objetos particulares e os conceitos), entram no que Kant deu o nome de “livre jogo das faculdades”. A reflexão é o ato por meio do qual este livre jogo pode ser representado e, dessa maneira, o sujeito toma consciência dele.

Quando o livre jogo das faculdades ocorre, nem a faculdade da imaginação nem a do entendimento são suficientes para providenciar a experiência adequada, e é essa insuficiência que faz com que nenhuma

faculdade prevaleça durante o juízo de gosto, ao mesmo tempo em que tomamos consciência dele pela reflexão. Daí a razão de o belo requerer articulação verbal e não ser uma experiência exclusivamente sensorial para Kant. O prazer que o belo nos proporciona (note que o belo não é o próprio prazer liberado) é o prazer que encontramos via o encontro, mediante a reflexão, da capacidade de julgar o que nos ocorre por meio do ajuizamento estético e do poder de comunicar isso às outras pessoas. Esse prazer advém de certa harmonia entre as faculdades que entram em livre jogo, pelo fato de nenhuma delas sobrepujar a outra na operação.

Essa harmonia nos conduz ao penúltimo ponto a considerar no pensamento kantiano sobre a beleza. Nos juízos de gosto acerca de obras de arte, ele pensa que nossas faculdades reconhecem o que chamou de forma de finalidade do objeto artístico. Nos objetos ou processos naturais – um pôr do sol, por exemplo – Kant argumenta que podemos reconhecer sua forma da finalidade, embora não saibamos qual é o “conteúdo” de tal finalidade. A suposição aqui parece ser a de que deus criou o mundo, e sua finalidade seria insondável pelos limites do conhecimento humano. Isso porque os limites do que podemos saber se encontram nas características do modo como todos os seres humanos percebem e conhecem o mundo. Por exemplo, quando vemos uma cadeira somos capazes de antever que a necessidade ou desejo humano de se acomodar confortavelmente guiou a produção do móvel e sabemos que ela não foi produzida ao acaso, ou seja, sabemos “para que”, com que finalidade ela foi produzida. Podemos conhecer tanto a forma quanto o conteúdo da “finalidade” da cadeira. Já nos objetos artísticos o reconhecimento desse conteúdo necessariamente desaparece da experiência; continuamos reconhecendo a forma da finalidade nos objetos artísticos, mas assim como os objetos da natureza (criados por deus) não somos capazes de reconhecer sua finalidade; tal reconhecimento é um dos responsáveis pela liberação do prazer que está envolvido na experiência do belo.

Por fim, para encerrar essa breve, mas intrincada e imperfeita apresentação, há a introdução de uma explicação daquela exigência de concordância universal nos juízos estéticos de gosto. Kant afirma que há algo como um *sensus communis*, isto é, um “sentido comum”; ele nos diz que se quisermos explicar por que há uma concordância universal sobre o belo e sua comunicabilidade, precisamos admitir que há uma capacidade de partilhar com

todas as pessoas esse sentido.

Dada a enormidade da análise de Kant sobre o belo, seu pensamento foi criticado por diversas vias – o qual talvez tenhamos apenas arranhado aqui. Um ponto bastante questionável parece ser o que de que, para haver uma experiência do belo, precisa haver o que chama de *desinteresse*. Para citar apenas alguns casos paradigmáticos, há obras literárias que exigem interesse (no sentido kantiano), pois fazem referências externas que estão na base do que são e em sua experiência plena. *Os Irmãos Karamazov*, por exemplo, supõe uma série de julgamentos morais, para não falar nas identificações pessoais que envolvem a encenação de tragédias e dramas. Nelas, o envolvimento moral e político está presente na própria experiência da beleza, poder-se-ia argumentar. Em *Memórias Póstumas de Brás Cubas*, por exemplo, o personagem-narrador menciona que estudou em Coimbra, mas que seus estudos não passaram de aparências, deles retirando apenas um “verniz”. Se essa referência e alusão política e moral, duas características da composição que contariam como interesse, ficassem de fora, dificilmente poderíamos considerar a obra bela durante sua experiência – o que não parece ser o caso para muitas pessoas. Mesmo passagens religiosas, como a chamada Paixão de Jesus Cristo, narrada nos Evangelhos, não poderiam ser origem da experiência de beleza – como muitos alegam – já que envolvem também referências externas indispensáveis: geográficas, políticas e a ao menos uma divindade.

Também há quem prefira enfatizar algumas condições históricas – sociológicas e culturais – na quais Kant formulou sua colossal teoria. Durante toda a segunda metade do século XVIII, sobretudo na parte ocidental da Europa, sabemos que a arte era praticada de acordo com normas bastante rígidas, reguladas na maior parte das vezes por instituições não oficiais, mas também por elas. A escrita de poemas obedecia rigorosos padrões métricos; a pintura, rigorosos padrões representacionais, assim como a escultura. A inovação artística estava também presente; acontecia muito mais lentamente, no entanto, e requeria uma aceitação do seletor público apreciador de tais obras. Exigir a não familiaridade com os procedimentos, como Kant parece pensar que a experiência do belo exige, pode ser apenas uma universalização apressada de um tipo específico de instituição cultural, composta por pessoas pertencentes a poucos grupos, e uma rejeição antecipada de outras. A universalidade e validade das fórmulas artísticas então consagradas pode ser vista também como certa

cegueira para as várias seleções (conscientes ou não) de quais tipos de obra dão origem a experiências do belo – assim como a outros valores associados à arte, como o sublime, a liberdade, o desenvolvimento da sensibilidade e tantos outros.

5. Que valor a arte tem?

É um ponto tão pacífico que a arte é importante para as pessoas que pode parecer desinteressante perguntar que valor ela tem. Todos os dias milhões de pessoas mundo afora visitam museus e galerias, vão ao cinema, leem romances e poemas, ouvem música, sem contar os milhões de artistas que devem estar nesse exato momento trabalhando em alguma nova obra. Diante de tanto interesse por elas, independente do que sejam exatamente, seria no mínimo extravagante sustentar que elas não têm valor ou que muitas pessoas não as valorizam de alguma maneira. Mas muita gente ainda se pergunta: que valor é esse? Por que valorizamos tanto esses objetos?

Parece ser tão óbvio que a arte possui valor – e um alto valor – que “obra de arte” pode ser usada como uma expressão que confere alto valor a um objeto. Quando num programa de culinária alguém diz “uau, a apresentação desse prato está uma obra de arte!”, a expressão é claramente usada num sentido honorífico.

Esse sentido honorífico, no entanto, pode trazer certa confusão à arte. Pois uma coisa é se referir às obras de arte visando classificá-las, outra muito diferente é utilizar a expressão para conferir valor a um objeto, ainda que em muitos casos a distinção não apareça ou apareça de maneira nebulosa.

É bem provável que se pudéssemos perguntar a cada uma dessas pessoas por que elas se interessam por obras de arte, a ponto de dedicarem seu tempo e recursos a elas, encontraríamos as mais diversas justificativas. Há obras de arte caríssimas, cujo valor econômico é desconcertante e que podem servir para formar patrimônio (e para atividades menos nobres que essa). Muitos tenderiam a dizer que seu valor está no prazer e diversão que proporcionam; outros a apreciam por questões religiosas ou para celebrar um momento importante da vida, como um aniversário; outros que buscam conhecer mais a si mesmos, o mundo ao seu redor, o passado e talvez imaginar o futuro; é provável também que haja quem queira investigar as profundezas psicológicas dos seres

humanos por meio da arte e quem queira apenas se tornar mais polido, ilustrado, aprendendo novas expressões e idiomas.

Vimos algumas tentativas de definição do que é uma obra de arte; o sentido honorífico do emprego do termo, no entanto, aponta para outro aspecto das obras: o seu valor. Olhando para essas formas de valorização da arte, muitos filósofos se perguntaram: o que de *especificamente artístico* possui valor ou é valorizado nas obras de arte? Para dar um exemplo: podemos aprender novos idiomas por meio de um romance numa língua desconhecida, mas esse é um valor que não é *específico* de um romance; uma boa gramática ou um manual pedagógico do idioma teriam mais valor a esse respeito. A seguir veremos alguns tipos de resposta sobre a especificidade do valor das obras de arte.

5.1. Hedonismo

Da perspectiva de um hedonista, se saímos de casa para assistir a uma comédia no teatro ou a um filme de terror, estamos todos primariamente buscando uma experiência de prazer, mesmo que outras experiências relevantes estejam envolvidas no percurso. Podemos nem ter o prazer em mente quando saímos de casa, mas todas essas experiências que temos em vista seriam redutíveis ao prazer. Pois o prazer seria a única coisa desejável em si mesma. Por exemplo, o que explica o valor da atividade de um ginasta se preparar por meio de treinos bastante desconfortáveis seria a experiência prazerosa de vencer a competição ou mesmo o prazer associado ao aprimoramento constante de sua técnica. É o prazer o que interessa em última instância.

O valor da arte, do ponto de vista hedonista, reside no prazer provocado pelas obras de arte. Sendo assim, se visitamos o Museu do Louvre em busca de alguns instantes de observação direta da *Mona Lisa*, de Leonardo da Vinci, é porque estamos em busca de certo prazer que essa atividade nos traria.

'Prazer' geralmente é associado à experiência de satisfação com alguma atividade, como uma conversa amigável, um bom jantar, a audição de uma boa música, o uso de substâncias como álcool e a cafeína etc. A experiência de prazer relacionada à arte geralmente está associada ao entretenimento, ao divertimento, à contemplação e assim por diante. Tomemos essas indicações sobre o que é o prazer associado à arte como suficientes, pelo

menos no contexto dessa breve apresentação.

Essa visão do valor da arte tem um apelo muito forte atualmente, dado o papel que a enorme indústria do entretenimento tem mundo afora. E também porque parece difícil negar que o prazer seja um componente fundamental da experiência artística. Caso visitar uma galeria não fosse prazeroso em algum sentido, pareceria estranho o fato de as pessoas voltarem regularmente ao mesmo local para apreciar obras de arte. Uma pessoa que quer se entreter com um filme de comédia dificilmente discordaria que essa experiência proporciona algum prazer, como talvez dissesse que é precisamente isso que está buscando. O mesmo poderia ser dito sobre a contemplação de um poema num sarau literário ou a audição de uma música barroca.

Podemos resumir a posição hedonista no seguinte argumento:

1. A arte é capaz de produzir prazer.
2. O prazer é a fonte última de valor.
3. Portanto, o valor da arte reside na sua capacidade de produzir prazer.

As premissas 1 e 2 parecem plausíveis, de modo que a conclusão 3 parece inevitável. Mas será mesmo?

Examinemos a premissa 1. Em que sentido a arte é capaz de produzir prazer? Certamente o dono de uma galeria e um artista de rua sentem prazer quando conseguem encontrar um comprador para suas obras de arte, mas não parece ser esse tipo de prazer que está em jogo aqui. Para um hedonista, o valor da arte tem origem na experiência com as qualidades das próprias obras, tais como o andamento da música, o ritmo e a força das metáforas de um poema, o detalhamento de uma escultura e assim por diante. São essas qualidades que propiciam o prazer que eles têm em mente. Mas o que é mais precisamente o prazer que as obras de arte propiciam pode ser entendido de maneiras diferentes pelos hedonistas.

Há pelo menos duas formas de entendê-lo. A primeira consiste em apresentar esse prazer como idêntico a outros tipos de prazer – chamemos esse ‘hedonismo B’ (em homenagem ao filósofo inglês Jeremy Bentham, um influente hedonista). Por exemplo, ouvir uma sonata ou ler *A Divina Comédia*, de Dante, produziriam o mesmo tipo de prazer que um glutão sente ao comer e beber

exageradamente. Pode haver um grande prazer envolvido em ambas as atividades e diferenças sobre qual delas proporciona mais prazer; no entanto, esses prazeres seriam qualitativamente iguais, diferindo apenas em fatores como quantidade, duração e intensidade. Outra versão do hedonismo – chamemos ‘hedonismo M’, em homenagem ao também inglês, e sucessor intelectual, J. S. Mill – sustenta que existem tipos de prazer qualitativamente distintos; de acordo com essa forma de ver o prazer, não seria razoável igualar o prazer proporcionado pela leitura de *A Divina Comédia* com uma quantidade de prazer despertada pela ingestão de várias barras de chocolate. O prazer oriundo do grande poema italiano seria não apenas qualitativamente distinto, como superior ao prazer do chocolate. Todo aquele que pensa ser o hedonismo uma boa explicação do valor das obras de arte precisa se debruçar sobre essa última distinção. Qualquer que seja a versão escolhida, o hedonismo não parece sair comprometido dela.

Intuitivamente muitos diriam que é óbvia a superioridade do prazer da leitura da *Divina Comédia* em relação ao prazer da ingestão de barras de chocolate, mas as coisas não são tão simples quanto parece. Pois de que modo poderemos constatar essa superioridade? J. S. Mill pensava que um “juiz competente” depois de provar os dois tipos de prazer daria o veredicto de que leitura da *Divina Comédia* é superior. Como ele chegaria a tal conclusão? Esse “juiz” possui alguma sensibilidade às propriedades estéticas dos objetos da qual carecem pessoas comuns? Não é possível que uma pessoa seja sensível às propriedades estéticas e ainda assim prefira o prazer do chocolate? Mill parecia pensar que não. Ele tinha em mente que para ser um juiz competente uma pessoa teria que ter experiências as mais diversas e que naturalmente ao repetir tais experiências ficasse claro a superioridade das experiências tais da arte e da ciência, por exemplo.

Independente da noção de prazer que se assuma, há contraexemplos claros à conclusão do argumento hedonista. Considere a obra *Bogey Ball*, de James R. Ford. Esse artista plástico passou algum tempo coletando muco do próprio nariz para formar uma bolinha de cerca de dois centímetros de diâmetro. No fim das contas, conseguiu vender sua obra por 20 mil dólares, o que indica que ela foi valorizada pelo mundo da arte. Dificilmente as pessoas costumam associar o nojo a uma experiência de prazer. Esse é um caso de obra de arte valorizada por seus apreciadores e que recebeu elogios razoáveis de críticos e

visitantes de museus, mas que não parece envolver prazer nenhum. Se é assim, então a conclusão do argumento é falsa, já que há pelo menos um caso de obra de arte que possui valor e que não provoca prazer. Uma vez que o argumento é válido, entretanto, invariavelmente teremos de rejeitar pelo uma das suas premissas. Dada a grande plausibilidade da premissa 1, só nos resta abandonar a premissa 2, que é exatamente o cerne da tese hedonista. Note que o contraexemplo compromete a tese hedonista de que é o prazer o responsável pelo valor da arte, embora possa não comprometer a explicação hedonista de outros tipos de valor, como por exemplo, o valor moral.

O hedonista, contudo, poderia tentar escapar à objeção de pelo menos duas formas. A primeira, mais radical, seria simplesmente negar que obras como *Bogey Ball* têm qualquer valor, uma vez que não proporcionam prazer. Essa seria uma forma estranha de tentar resolver a questão, pois terá que assumir que muitas pessoas estão enganadas a respeito do valor que dão a essa obra e teria o trabalho de ter de explicar por quê. Uma segunda, mais moderada, poderia tentar argumentar que o prazer que uma obra como *Bogey Ball* propicia não é imediato. O prazer que ela propiciaria aparece depois que obra é apreciada, quando se entende os desafios conceituais que ela propõe ao apreciador e a tomada de consciência intensa de certos estados corporais; a possível sensação de nojo seria compensada pelo prazer ulterior de compreensão da obra, por exemplo.

Não está claro se, ou em que medida, essas duas saídas resolvem os desafios colocados ao hedonismo. Serão essas saídas razoáveis? Especialmente a última: é possível entender que qualquer experiência artística, mesmo aquelas que não envolvam prazer imediato, como tendo valor apenas em virtude do prazer que proporcionam em última instância?

5.2. Cognitivismo

Em estética, o cognitivismo é outra família de teorias filosóficas que procura explicar o valor especificamente artístico das obras de arte. Todo cognitivista sustenta que o valor da arte se deve a sua capacidade de proporcionar conhecimento aos seus apreciadores, embora haja muita divergência sobre o que isso significa precisamente. Ler o *Dom Quixote*, de Miguel de Cervantes, pode nos fornecer conhecimento histórico ou linguístico,

por exemplo, mas esses conhecimentos são laterais, já que não decorrem das propriedades estéticas da obra. O cognitivista não pensa que obras de arte forneçam conhecimento simplesmente porque podem ser usadas como evidência histórica ou linguística. Há outras formas de obter esse tipo de evidência a partir de objetos não artísticos. O que torna o cognitivismo uma tese acerca do valor é a suposta existência de um tipo de conhecimento que decorre única e exclusivamente das propriedades estéticas das obras de arte. Ainda que um romance seja composto totalmente por proposições falsas, para um cognitivista é possível obter conhecimento a partir dele.

As várias versões do cognitivismo convergem para a conjunção de duas afirmações:

- a. Todas as obras de arte nos fornecem conhecimento.
- b. Esse conhecimento confere valor à arte.

Observe que a tese parece demasiado ampla. Parece difícil sustentar que todas as obras de arte, inclusas *A Fonte* e as telas de Pollock, fornecem conhecimento. Por outro lado, a literatura, a pintura e a escultura representacional parecem fortes candidatos a fontes de conhecimento de algum tipo. Mas que tipo de conhecimento decorreria das propriedades estéticas de uma obra?

Quadro 4: Conhecimento proposicional e não-proposicional

A forma mais tradicional de abordar a questão consiste em separar ao menos três sentidos nos quais o termo conhecimento é utilizado. O primeiro sentido é chamado de *proposicional* – também conhecido como “saber que” – é expresso em frases como “eu sei que o céu é azul”, “eu sei que a água ferve a 100° C” e outras do mesmo tipo. O importante aqui é a ideia de que sabemos alguma coisa sobre o mundo. Já quando dizemos “minha irmã sabe andar de bicicleta”, utilizamos o termo ‘saber’ em outro sentido; no sentido de que minha irmã possui a habilidade ou a capacidade de andar de bicicleta. O terceiro sentido que o termo assume aparece em frases como “eu conheço o Pão de Açúcar, no Rio de Janeiro” ou “eu conhecia aquele cantor pessoalmente”; ambas querem dizer que tivemos algum tipo de contato ou temos familiaridade com certas coisas. O primeiro sentido é chamado de proposicional. Os outros dois são classificados como não-proposicionais.

A diferença entre esses dois tipos de conhecimento está no que chamamos de “justificação”. Quando dizemos que uma pessoa qualquer sabe alguma coisa sobre o mundo – que o céu é azul, por exemplo – dizemos que há pelo menos três outras condições para que o termo ‘saber’ ou ‘conhecimento’ esteja sendo utilizado apropriadamente. A pessoa que profere a proposição precisa acreditar que essa frase é verdadeira: não há como alguém dizer que sabe que o céu é azul sem acreditar nisso. Outra condição é que a frase seja verdadeira: não há como alguém propriamente saber que há quadrados que são círculos ou que o céu é amarelo, pois em seus sentidos literais ambas as proposições são falsas. A terceira e mais importante é a justificação. Uma pessoa pode, por exemplo, formar a crença de que os números da loteria serão 01 – 09 – 36 – 25 – 60 em virtude de um sonho com eles e os números sorteados serem exatamente esses. Essa pessoa acreditou numa proposição que se revelou verdadeira, mas parece razoável não aceitar que ela de fato sabia quais seriam os números sorteados, pois as razões que ofereceu não pareciam sustentar apropriadamente essa crença. De modo geral, podemos dizer que:

S (uma pessoa qualquer) sabe que p (onde p é uma proposição qualquer sobre o mundo) sse (i) S acredita que p; (ii) p é uma proposição verdadeira; (iii) S tem boas razões para sustentar sua crença em p.

Para maiores detalhes, ver o capítulo sobre Epistemologia neste volume.

Começemos pelo conhecimento proposicional. Que tipo de proposição exatamente o sujeito poderia extrair de uma obra de arte? Voltemos ao *Dom Quixote*: após a leitura de alguns capítulos, podemos vir a *saber que* Amadis de Gaula era uma influente novela da cavalaria no tempo de Cervantes. Um historiador certamente poderia utilizar o romance para justificar sua crença nessa proposição. Mas um historiador pode justificar tal crença de outras formas, recorrendo a um exemplar datado da novela de cavalaria, por exemplo. Esse tipo de conhecimento nada tem a ver com as qualidades artísticas de *Dom Quixote*. Em geral, parece difícil que obras de arte possam oferecer qualquer tipo de conhecimento factual. Se é assim, que tipo de conhecimento poderia ser obtido a partir das qualidades artísticas de *Dom Quixote*?

Há cognitivistas que sustentam que obras de arte podem fornecer conhecimento de contrafactuais, isto é, conhecimento sobre como as coisas poderiam ter sido, caso o mundo fosse de determinada maneira. Em *Mad Max*, somos levados a um universo ficcional onde há extrema escassez de recursos,

num mundo onde o ambiente está totalmente degradado. O modo pela qual a narrativa é apresentada – o roteiro, as tomadas, a atuação dos atores, as cores utilizadas, o enquadramento etc. – ilumina vários aspectos da natureza humana, o que nos permite justificar a crença de como tenderíamos a nos comportar numa situação real semelhante.

Outras versões do cognitivismo sustentam que o conhecimento obtido é do tipo não-proposicional. Pense num romance como o *1984* de George Orwell. O protagonista da história, Winston Smith, quer se encontrar com sua amante Júlia; no entanto, há inúmeras câmeras capazes de monitorar seus passos: as teletelas. Se for pego, sofrerá sérias consequências, podendo ser entregue ao Ministério do Amor; isso o obriga a elaborar planos cuidadosos para não ser notado. Imaginando uma situação como a narrada, o leitor pode desenvolver uma maneira mais prudente de raciocinar, buscando antever os obstáculos aos fins que pretende alcançar – o chamado raciocínio prudencial. Essa capacidade de ser mais prudente seria um tipo de conhecimento não-proposicional, análogo à capacidade ou habilidade de andar de bicicleta, desencadeado pela leitura de *1984*. Esse caso permite aumentar nosso “conhecimento”, em sentido não proposicional, e esses exemplos poderiam ser multiplicados ou adaptados aos montes. O importante é que certas qualidades das obras de arte forneçam conhecimento não-proposicional.

É difícil negar que possamos obter conhecimento contrafactual e que podemos exercitar o raciocínio prudencial. Em ambos os casos, precisamos usar a imaginação para tentar conceber cenários alternativos ao que estamos vivendo. E não há outra maneira de fazer isso a não ser por meio de uma narrativa. Parece claro, contudo, que inventar uma narrativa não é suficiente para que tenhamos uma obra de arte. Um testemunho de um amigo pode nos ajudar a conceber tais cenários e nem por isso é uma obra de arte. É um tanto plausível pensarmos que um relato não artístico dos encontros de Júlia e Winston Smith ou da briga dos habitantes da cidadela de Imortal Joe, por água, fossem igualmente eficazes no treino do raciocínio prudencial ou na obtenção de conclusões sobre aspectos da natureza humana. As propriedades artísticas das narrativas originais são completamente acessórias; certamente contribuem para a leitura se tornar mais viva e estimulante, mas não são indispensáveis para obtenção de tais conhecimentos. A imaginação é suficiente. Portanto, o conhecimento supostamente obtido dessas obras de arte não decorre

exclusivamente de suas propriedades artísticas. Por conseguinte, o cognitivista se equivoca ao atribuir às propriedades artísticas dessas obras um papel central na obtenção de conhecimento.

5.3. Formalismo

O hedonismo e o cognitivismo têm em comum o fato de serem teorias que explicam o valor da arte em termos instrumentais, isto é, a arte adquire valor na medida em que é um meio para o alcance de determinado fim. No caso do hedonismo, o prazer; e no caso do cognitivismo, o conhecimento. Por isso tais teorias são classificadas como teorias “instrumentalistas” do valor da arte.

Podemos, contudo, pensar na arte como possuidora de valor em si mesma – como tendo valor intrínseco. A ideia é a de que não precisamos buscar o valor da arte em algo externo à própria arte. Uma famosa expressão de origem francesa dá corpo a essa intuição: a arte pela arte. Tem a intuição de que a experiência da arte é totalmente desinteressada, livre de qualquer necessidade prática ou teórica.

Os formalistas sustentam que o valor da arte não pode ser instrumental. Alegam que se o valor da arte for instrumental, então seu valor não seria propriamente artístico, e sim meio para a realização de outro valor qualquer. Considere a proposta hedonista. Se for o prazer o que explica o valor da arte, seria o prazer o que teria valor, não a obra de arte, que assim estaria na posição de “mero” instrumento. O mesmo se aplica para outros valores instrumentais tais como o econômico, o documental, o cognitivo etc.

A mais importante teoria a defender que a arte tem valor intrínseco é o formalismo, segundo o qual o valor da arte depende única e exclusivamente de suas propriedades formais. O formalismo não nega que as obras de arte possam ter certos valores instrumentais, mas que esses valores são secundários em relação ao valor especificamente artístico das obras de arte. Elas podem até representar algo, podem nos dar prazer e assim por diante, mas seu valor enquanto arte reside em suas características formais e suas estruturas internas, de modo inteiramente autônomo a qualquer aspecto exterior.

Um poderoso argumento invoca casos de obras de arte, talvez um ramo inteiro delas, cujas propriedades estéticas não parecem representar ou expressar o que quer que seja. Eduard Hanslick, um filósofo, músico e escritor

austríaco do século XIX, apresentou uma teoria formalista que visava a defender que o valor da música absoluta – ou seja, da música puramente instrumental – se devia a suas propriedades formais e a nada mais. Para isso, apresentou dois argumentos centrais: um negativo e um positivo. Em primeiro lugar, buscava convencer racionalmente que a música absoluta *não* possui valor instrumental. Em segundo, buscou convencer que o valor da música absoluta reside em suas estruturas e formas que lhes são próprias.

Examinemos a seguir o argumento negativo, a partir do qual podemos avaliar a posição geral de Hanslick:

1. Se a música tem valor instrumental, então se remete a algo externo à música.
1. A música absoluta não se remete a algo externo à música.
2. Logo, música não tem valor instrumental.

Passo a passo, examinemos o argumento começando pela primeira premissa; ela nos diz que se a arte tiver algum valor instrumental, como vimos, for meio para alcançar algum fim valoroso, então ela precisa “se remeter a algo externo à música”. Isso quer dizer que fazer referência a algo externo é uma condição necessária para que a música tenha valor instrumental; pois, como também já vimos, o valor instrumental é sempre dependente de outra coisa que tenha valor, como o prazer.

A música remete a algo externo na medida em que “representa” ou “expressa” algum conteúdo extramusical. A ideia que esses termos sugerem podem ser bastante complexas, mas de modo geral querem dizer algo como “simbolizar”, “descrever”, “imitar”, “substituir” um objeto qualquer. Do mesmo modo como as letras “b” e “r” se remetem aos fonemas “bê” e “érre” no português brasileiro, uma pintura que contém uma imagem semelhante em vários aspectos a Diego Maradona – traços do seu rosto, sua estatura, uma camisa da seleção argentina de futebol etc. – “representa” a pessoa de Diego Maradona. De acordo com Hanslick, a música não pode fazer nada desse tipo, pois não existe um conjunto de convenções sonoras que permitam representar um objeto no mundo por meio de sons puramente instrumentais. Nem mesmo as emoções poderiam ser representadas, pois nada indica que o andamento de

uma música represente a alegria ou a tristeza, embora possam ser fortuitamente associadas a tais emoções. As únicas coisas que a música poderia representar desse ponto de vista são as ideias musicais, tais como a mudança de ritmo, a altura do som, as variações e outras mais; mas note que nenhum objeto ou emoção que estejam “fora” da música.

Alguns poderiam argumentar que na *Abertura 1812*, de Tchaikovsky, por exemplo, os tiros de canhões poderiam representar coisas como uma guerra, momento intenso de conflito ou uma mescla sugestiva de coisas assim. Mas esse seria um caso, dentre alguns outros, em que uma representação um tanto confusa seria feita por meio da semelhança entre os sons e não por meio de convenções musicais.

Tendo tomada como sustentada a tese segundo a qual a música absoluta não representa ou expressa algo, Hanslick passa a sustentar que o conteúdo da música está exclusivamente nas “formas sonoras em movimento”. Não podendo ser meio para mais nada, seu valor seria totalmente autônomo. Seu argumento positivo sobre o valor da arte pode ser formulado aproximadamente por:

- Se a música não tem valor instrumental, seu valor é intrínseco.

 1. A música não tem valor instrumental.
 2. Logo, seu valor é intrínseco.

Observe que esse argumento depende da força do primeiro argumento. Note também que a conclusão do argumento negativo constitui a premissa 2 do argumento positivo. A premissa 1 do argumento positivo é bastante simples; é apenas a constatação de que, se a música não tem valor instrumental, tem valor intrínseco (supondo, é claro, que a música tenha valor). Assim, se o primeiro argumento for bom, o segundo também será. De tal sorte que se houver algum problema nas premissas que sustentam a conclusão do primeiro argumento, o segundo também teria que ser abandonado.

É importante prestar atenção ao fato de que o argumento de Hanslick diz respeito apenas à música, e o nosso interesse nessa seção se dirige à arte em geral. De modo que para defendermos o formalismo mais geral teríamos que generalizar o argumento de Hanslick em direção às demais artes. Fica ao leitor o exercício de tal generalização.

Concentremo-nos, agora, no argumento negativo. Para avaliá-lo, temos que levar em conta dois pontos: a) representar algo é condição necessária para a música ter valor instrumental?; b) a música absoluta realmente não representa nada além de ideias musicais?

Talvez Hanslick tenha razão em pensar que a insistência num suposto conteúdo representacional de uma peça musical desvia a nossa atenção da genuína fonte de valor da música. Disso não se segue, todavia, que o valor instrumental tenha qualquer relação íntima com a representação. Caso o hedonista esteja correto, o valor da música depende exclusivamente do prazer por ela provocado. E aqui o hedonista não é específico sobre qual é exatamente a origem causal do prazer provocado pelas obras de arte. É razoável pensar que sejam tanto o padrão físico do som quanto seu suposto conteúdo representacional. O que importa é que a música provoque prazer. Portanto é perfeitamente concebível que a música tenha valor instrumental, mesmo não tendo conteúdo representacional extramusical. Se esse ponto ficar estabelecido, então a condicional que forma a primeira premissa do argumento negativo de Hanslick é falsa.

O ponto (b) se refere à segunda premissa, que nos diz que a música não se remete a algo externo a ela mesma. Ora, não há consenso de que os sons não possam representar qualquer coisa fora das próprias ideias musicais (tais como andamento, altura e variações). Alguns argumentam, por exemplo, que a música é capaz de representar emoções. De maneira semelhante à pintura, que representa certos objetos conforme as nossas experiências visuais, permitindo a constatação da representação pela semelhança visual direta, por simbologia baseada em outras convenções – como a de que um coração pode representar a ideia do amor –, a música seria capaz de representar emoções. O próprio Hanslick reconhece que existe certa semelhança entre certas propriedades das emoções e as propriedades dinâmicas da música. O nível de tensão criado por uma música, bem como sua intensidade, variações e velocidade poderiam representar certas emoções definidas.

Disso tudo se segue que o formalismo como um todo é implausível? De maneira alguma. Primeiro porque aventamos apenas uma versão dele. O principal é notar que, tal como formulada aqui, essa teoria teria que, no mínimo, ser reformada, de modo a responder aos questionamentos levantados.

7. Questões para revisão

1. O que qualquer teoria da arte precisa oferecer? Há alguma exceção?
2. Que contraexemplos seriam difíceis de acomodar numa teoria da imitação? Por quê?
3. O que seria um “conceito aberto”?
4. Qual seria a diferença entre a beleza “subjetiva” e “objetiva”?
5. Como Hume espera encontrar o padrão do gosto?
6. O que é uma teoria do valor da arte?
7. O que é uma teoria instrumentalista e não-instrumentalista do valor da arte?

8. Para pensar

1. Seria possível reformular a teoria da imitação de modo a resolver os problemas levantados?
2. É possível ser ao mesmo tempo hedonista e cognitivista?
3. Uma música puramente instrumental poderia fornecer conhecimento?
4. Tente imaginar como seria uma teoria não-instrumentalista do valor da arte não formalista.

9. Indicações de leitura

Você deve ter notado que todas as propostas teóricas precisam de refinamento, caso queiram se apresentar como boas candidatas a oferecer uma resposta para essas questões filosóficas. Caso você queria começar a se aprofundar mais no tema, seguem algumas indicações de leituras, que, por sua vez, vão indicar a você outras mais. Se esse texto foi capaz de ajudá-lo ou ajudá-la a entrar nessa conversa, cumpriu o seu objetivo.

9.1 Sobre a definição de arte:

ALMEIDA, A. A Definição da Arte O essencial. Lisboa: Plátano Editora, 2019.

DICKIE, G. Introdução à Estética. Lisboa: Bizâncio, 2008.

OREY, C. D. O que É a Arte? A Perspectiva Analítica. Edição: 1ª ed. Lisboa: Dinalivro, 2007.

9.2 Sobre os conceitos de gosto e beleza:

DUARTE, R. **O belo autônomo - Textos clássicos de estética**. Edição: 2 ed. Belo Horizonte: Autêntica, 2012.

GRAHAM, G. **Filosofia das Artes**. Lisboa: Edições 70, 2001.

KIVY, P. (ED.). **Estética: Fundamentos e Questões de Filosofia da Arte**. Edição: 1ª ed. São Paulo: Paulus Editora, 2008.

NUNES, B. **Introdução à filosofia da arte**. Rio de Janeiro: Edições Loyola, 2016.

ROOKMAAKER, H. R. **Filosofia E Estética**. Brasília: Monergismo, 2017.

9.3 Sobre o valor da arte:

ALMEIDA, A. **O valor cognitivo da arte**. Lisboa: Universidade de Lisboa, 2005.

9.4. Sobre outros tópicos:

ALMEIDA, A. **A Definição da Arte O essencial**. Lisboa: Plátano Editora, 2019.

WARBURTON, N. **The Art Question**. Edição: 1ª ed. London; New York: Routledge, 2002.

WIETZ, M. **O papel da teoria na estética**. Trad.: Célia Teixeira. Critica na Rede, 2004.

6

Filosofia da Ciência

Tiago Luís Teixeira de Oliveira

Introdução: Filosofia, Ciência e Filosofia da Ciência

Seria muito difícil imaginar o modo de vida que atualmente possuímos sem a contribuição das ciências. Geladeiras, computadores, lâmpadas elétricas, impressoras, forno de micro-ondas, carros, aviões e outros sistemas de transporte, remédios, tratamentos médicos, vacinas, cirurgias, cosméticos e materiais de higiene pessoal e limpeza são apenas alguns exemplos de aplicação do conhecimento científico ao nosso cotidiano. É também muito difícil imaginar um empreendimento tão bem-sucedido quanto a ciência. Se compararmos o progresso feito em campos do conhecimento como a medicina, a genética e a própria física teórica em relação ao estado dessas mesmas áreas há dois séculos, compreende-se o porquê de tal avaliação sobre a ciência.

Houve, porém, um tempo (pelo menos até o século XIX de nossa era) em que aquilo que hoje chamamos de ciências naturais compunha um ramo da filosofia conhecido como “filosofia da natureza” ou “filosofia natural”. Um dos marcos do pensamento científico, o texto em que Isaac Newton apresenta as suas leis do movimento, tem o título de *Princípios Matemáticos de Filosofia Natural*. Não é, portanto, ao acaso que ciência e filosofia guardem certas similaridades (ambas constituem uma tentativa de buscar, com rigor, respostas corretas aos problemas que circunscrevem seus respectivos domínios). O filósofo britânico Bertrand Russell (1872-1970) em seu *Os problemas da filosofia* (publicado pela primeira vez em 1912) chegou a sugerir que a aparente falta de progresso na filosofia devia-se ao fato de que, tão logo um problema alcançasse um tratamento metodologicamente consensual, tal problema se tornaria científico. Na visão de Russell, uma nova ciência incorporaria o método recém-criado e os

resultados dele advindos e, dessa forma, as questões que continuassem intratáveis através de um método científico continuariam a fazer parte do escopo da filosofia. Embora o relato de Russell pareça acomodar bem o fato de que a física, a biologia ou a psicologia nasceram de reflexões filosóficas e depois se tornaram autônomas em relação à filosofia, tal proposta pode dar a entender que em um último estágio, as lacunas deixadas para a atividade filosófica ficariam cada vez menores até que só sobrariam as ciências. Sem dúvidas, não era isso que Russell tinha em mente. O filósofo inglês sugeriu que algumas questões estariam fora do escopo científico, por sua própria natureza, e, mesmo aparentemente insolúveis, deveriam ser investigadas filosoficamente, sob o risco de perdermos muitos benefícios intelectuais, os quais tornam a vida humana verdadeiramente valiosa e interessante. Esse tipo de abordagem laudatória da ciência muitas vezes omite um dado que talvez só se justifique pela diferença de importância que nossa sociedade confere à ciência e à filosofia (em favor da primeira e contra a segunda). O dado que tenho em mente é o de que não só problemas científicos emergem da reflexão filosófica cuidadosa, mas também a teorização e a prática científica criam novos problemas filosóficos. Tais problemas advindos da ciência demandam, para serem investigados adequadamente, que cientistas possuam alguma iniciação filosófica ou procurem a colaboração de filósofos da ciência.

Há uma série de perguntas legítimas, que podem ser feitas a partir do conhecimento científico, que não parecem muito passíveis de resposta dentro do âmbito da ciência em causa. A título de exemplificação, muitos pesquisadores admitem que determinados fenômenos causam outros fenômenos (uma infecção viral causa aumento de linfócitos no sangue), mas o que é uma causa? Outros tantos pesquisadores criam modelos (como o modelo atômico e o da dupla hélice de DNA), mas até que ponto e em que grau tais modelos estão associados à realidade? Algumas ciências dependem sobremaneira de teorias que supõem partículas fundamentais inacessíveis a olho nu, mas tais partículas são entidades que compõem a realidade ou apenas construções lógicas? E como atestamos que elas existem, caso existam? O que significa quando se diz que uma teoria está confirmada? O que conta como confirmação? Supõe-se que o conhecimento exposto nos periódicos científicos é objetivo, no sentido de não depender de critérios obscuros ou subjetivos, arbitrariamente escolhidos pelos autores, mas de critérios públicos claros, segundo os quais qualquer investigador

da área poderia submeter o artigo a um escrutínio crítico. Mas isso significa que os autores da publicação não possuam valores e vieses que influenciaram seu trabalho? Será que a objetividade científica é um ideal realizável? Em última instância, o que é ciência e como ela se diferencia de outros saberes? Essas e outras questões similares compõem o âmbito da filosofia da ciência. Como se vê, não são problemas que um experimento rigorosamente conduzido poderia responder, e os cientistas, quando se dispõem a investigar tais questões, já não o fazem cientificamente, mas filosoficamente. Por essa razão, as discussões que competem à filosofia da ciência são de grande importância para filósofos e cientistas, ainda que filosofia e ciência tenham diferentes métodos e objetos.

A filosofia da ciência é o ramo da filosofia que se debruça a investigar a natureza do conhecimento científico, seus pressupostos lógicos e metodológicos; que interroga se, e em que medida, a realidade é desvelada pela ciência; que procura entender quais implicações não só metafísicas e epistêmicas, mas também éticas e políticas decorrem das ciências. A filosofia geral é tradicionalmente retratada como uma investigação sobre a natureza geral da realidade (metafísica), do conhecimento (epistemologia) e dos valores (axiologia). Pode-se, analogamente, dividir a filosofia da ciência em epistemologia da ciência, metafísica da ciência e axiologia da ciência. Por questão de espaço e escopo, tentarei expor os problemas investigados principalmente pelas duas primeiras áreas, mas irremediavelmente algumas discussões levarão a valores. A divisão por áreas é muito mais didática do que real, seja na filosofia geral, seja na filosofia da ciência. Isso significa que muitas questões metafísicas dependem de questões epistemológicas e vice-versa, muitas questões epistêmicas implicam ou são implicadas por valores e vice-versa. É preciso dizer também que cada disciplina científica proporciona problemas filosóficos suficientes para existência de uma filosofia para cada ciência particular. Existe, desse modo, a filosofia da física, a filosofia da psicologia, a filosofia da biologia, da neurociência, da lógica, da matemática, das ciências sociais etc. Não é objetivo deste texto introdutório abordar os problemas específicos da filosofia de cada ciência particular, mas o leitor terá a oportunidade de conhecer alguns desses problemas específicos nos capítulos deste volume dedicados à filosofia da física, da biologia, da economia e da mente (neste há questões de filosofia da psicologia).

1. Epistemologia da ciência

Os grandes problemas epistemológicos da ciência envolvem uma teoria do método científico, a distinção entre ciência e outros tipos de abordagens e saberes, e também uma tentativa de explicar a lógica subjacente às inferências científicas. Um dos maiores filósofos da ciência, o austríaco Karl Popper (1902-1994), resumiu as questões epistemológicas da ciência a dois problemas: a) o problema da demarcação e b) o problema da indução. Falaremos dos dois problemas de modo mais ou menos entrelaçado, já que muito da discussão sobre o que distingue a ciência empírica (isto é, experimental) de outras abordagens depende também de se entender em que medida as alegações das ciências estão justificadas e de que modo o raciocínio indutivo está implicado no método científico.

O problema da demarcação se traduz na tentativa de traçar uma linha divisória entre o que é e o que não é uma ciência experimental. A discussão sobre demarcação ganha maiores contornos e importância quando pessoas mal-intencionadas (ou mal-informadas, na melhor das hipóteses) procuram apresentar ideias, produtos e tratamentos de saúde à margem de qualquer padrão científico aceitável, como se fossem uma ciência séria ou técnicas administradas de acordo com uma ciência séria. Reivindicando o status de cientistas, homeopatas conseguem fazer com que seus produtos sejam incluídos em listas de medicamentos distribuídos por governos. E algumas denominações religiosas pedem para que teorias sobre um criador divino guiando os seres vivos segundo seus desígnios seja ensinada como alternativa à teoria da evolução das espécies nas aulas de ciências das escolas públicas. A própria filosofia, enquanto investigação, parece não dispor dos mesmos mecanismos metodológicos disponíveis para as ciências empíricas. Não por acaso, uma das primeiras tentativas de demarcação visava distinguir entre ciência e metafísica. Nesta primeira parte, procurarei apresentar três propostas de demarcação, a saber: verificacionismo, falsificacionismo e pluralismo historicista. Também procurarei expor o problema da indução em sua relação com as abordagens mencionadas e uma tentativa de resposta conhecida pelo nome de bayesianismo.

1.1 Verificacionismo e o problema da indução

O movimento filosófico que primeiro tentou avançar um critério de demarcação foi o Positivismo Lógico iniciado na década de 1920. De acordo com os pensadores desse movimento e muitos de seus herdeiros intelectuais, um enunciado científico seria caracterizado pela possibilidade de ser decomposto em proposições mais simples, de modo que só contariam como científicas aquelas cuja verdade ou falsidade pudesse ser estabelecida por um teste verificação empírica (isto é, pela observação) e aquelas que enunciassessem uma verdade lógica (uma tautologia, como as regras de dedução, por exemplo). Neste caso, os enunciados contendo termos supostamente referentes a entidades inobserváveis tais como elétrons, prótons, fótons, etc. não passariam de abreviações para procedimentos de mensuração. A ligação entre um termo e o procedimento observacional seria feita por “princípios de transposição”: o termo ‘elétron’ seria equivalente a ‘espectro obtido pela emissão de energia numa câmara de vapor’. Além de critério de demarcação, o Positivismo Lógico supunha que a verificabilidade também consistia num critério de significado, adotando um adágio que remonta a Wittgenstein (1889-1951), segundo o qual “o significado de uma proposição é o método de sua verificação”. Através do critério de verificabilidade seria possível reconhecer que uma proposição como ‘cloreto de sódio ao ser dissolvido em água conduz eletricidade’ é científica. O método de verificação é um teste simples no qual tenta-se, sem sucesso, conduzir eletricidade através de água destilada e, ao repetir o procedimento depois de adicionar sal à água, observar que a corrente é conduzida, acendendo uma lâmpada. A proposição ‘água destilada conduz eletricidade’ é falsa, porém sabemos como verificá-la, o que é suficiente para que tenha sentido, de acordo com o critério positivista. O mesmo não parece ser o caso de proposições metafísicas. Proposições envolvendo a vontade divina, a alma, o Ser, o Nada etc. careceriam de significado, de modo que sequer poderiam ser verdadeiras ou falsas. Rudolph Carnap (1891-1970), um dos grandes expoentes do positivismo lógico foi particularmente duro, ao analisar uma proposição enunciada pelo famoso filósofo alemão Martin Heidegger (1889-1976), de acordo com o qual “o Nada nadaifica”. A crítica positivista é a de que embora a frase pareça obedecer a regras gramaticais, não seria possível sequer compreender o que Heidegger queria dizer com ela, já que não se sabe o que a tornaria verdadeira ou falsa.

Um breve sobrevoo a toda a história da filosofia seria suficiente para perceber que boa parte da mesma seria nada mais que uma série de pseudoproblemas motivados por um palavrorio que mesmo os filósofos envolvidos eram incapazes de definir corretamente. Deste modo, para o positivista mais devoto, a filosofia só teria legitimidade como fornecedora de um método de análise lógica das proposições. O conhecimento factual do mundo só poderia ser estabelecido cientificamente.

Há muitíssimos problemas com a adoção da verificabilidade como critério de demarcação. Destacarei três. Em primeiro lugar, o critério de verificabilidade não é verificável. A frase ‘uma proposição é verdadeira se e somente se ela for verificável ou se ela for uma verdade lógica’ não é nem verdade lógica nem verificável. Para ser uma verdade lógica, seria necessário que sua negação fosse uma contradição. Imagine uma proposição $A =$ ‘Hitler levou a Alemanha à derrota’. Neste caso, não- $A =$ ‘Hitler não levou a Alemanha à derrota’. Indiferente a como foi a história, uma frase composta ‘ A ou não- A ’ é uma verdade lógica, pois é verdadeira em virtude de seu próprio significado, já que ou Hitler ganhou ou não ganhou. O mesmo ocorre com todas as frases que possuam a mesma forma lógica ‘ A ou não- A ’ para qualquer A : ‘chove ou não chove’, ‘estou certo ou não estou certo’, ‘é domingo ou não é domingo’. Negar uma frase da forma ‘ A ou não- A ’ é produzir uma contradição, como é o caso de dizer que ‘não- A e A ’. Não ser domingo e ser domingo é impossível quando se refere a um e o mesmo dia da semana. No caso da proposição que exprime o princípio de verificabilidade, não há nenhuma análise lógica que permita ser compreendida como uma verdade similar a ‘ A ou não- A ’. Ela parece ter a forma ‘ A se e somente se (B ou C)’ e só por convenção arbitrária pode-se transformar ‘ B ou C ’ em ‘ A ’. Por outro lado, também não há nenhum teste empírico de verificação que torne o princípio significativo sob seu próprio critério.

Em segundo lugar, a ideia de que termos teóricos para inobserváveis não passam de abreviações para procedimentos observacionais é estranha para o modo com os próprios cientistas se manifestam, ao usar tais termos. Estes sugerem muito mais estar lidando com realidades fundamentais, e não com meros construtos lógicos. E, mesmo que os cientistas estejam errados sobre esse aspecto, a tese profundamente empirista associada ao critério do positivismo lógico demanda que seja possível existir uma linguagem observacional totalmente livre de pressupostos teóricos. Em outras palavras, o

positivista precisa defender que a observação científica é independente de qualquer teoria científica. Isso parece contrariar o próprio uso de instrumentos de detecção, tais como microscópios, telescópios, *scanners*, reagentes químicos e computadores. Uma simples observação astronômica por telescópio ou luneta depende de assumirmos a correção de uma teoria óptica que permite a construção do aparelho (incluindo o modo como a luz se comporta, o modo como a imagem é transmitida e codificada, o modo como se espera que as lentes funcionem). Muitos filósofos da ciência consideram impossível dissociar a observação de seus pressupostos teóricos. Dois dos mais importantes defensores da dependência teórica são o austríaco Paul Feyerabend (1924-1994) e o estadunidense Thomas Kuhn (1922-1996), sobre os quais nos pronunciaremos mais tarde.

O terceiro problema associado ao critério de verificabilidade como demarcação entre ciência e outros discursos é a dependência de um modo de raciocinar que tem desafiado filósofos e cientistas inclinados à filosofia desde o século XVIII: a indução. O problema da indução consiste, basicamente, na incapacidade de filósofos e cientistas justificarem adequadamente o tipo de raciocínio que fundamenta a elaboração de leis gerais, de previsões e de explicações de fenômenos. Os cientistas parecem muito competentes em elaborar leis, dar explicações e fazer previsões, mas quando é preciso dizer o porquê de seu sucesso naquelas tarefas, é difícil apresentar boas razões para além da funcionalidade das teorias. Como poderíamos dizer, então, que a experiência confirma uma lei da física ou da química, se tudo o que sabemos é que os experimentos até agora conduzidos o fazem e que temos a esperança de que isso continue ocorrendo indefinidamente, ainda que não saibamos o porquê dessa esperança? Fundamentar a lógica indutiva é extremamente necessário para uma filosofia empirista da ciência (como é o caso do positivismo lógico), pois os raciocínios baseados em observações empíricas não possuem a mesma garantia que os argumentos chamados dedutivos são capazes de fornecer, quando válidos.

Provas matemáticas, teoremas da lógica e aplicação de fórmulas nas quais, de posse de certos valores, é possível descobrir um valor desconhecido através de cálculos e regras de inferência, são modos usuais de **raciocínio dedutivo**. Numa dedução válida, a verdade das premissas (dos enunciados com os quais iniciamos o raciocínio) garante a verdade da conclusão (aquilo que se

quer provar). Um exemplo de dedução é o silogismo disjuntivo, cuja forma mais básica é a existência de duas alternativas apenas (A ou B) e a afirmação de que uma delas não é o caso, resultando que só a outra opção é verdadeira. Assim, se só for possível chegar a uma ilha de barco ou a nado (A ou B), mas a pessoa que está na ilha não sabe nadar (não- B), deduz-se que ela chegou de barco (A). É em virtude da forma do raciocínio dedutivo que a conclusão decorre das premissas e, por isso, quaisquer que sejam A e B , o raciocínio seguinte é válido, isto é, se as premissas forem verdadeiras, sua conclusão também será:

A ou B	(premissa)
Não- B	(premissa)
Logo, A	(conclusão)

Os argumentos indutivos, diferentemente dos dedutivos, são aqueles em que a conclusão diz mais do que as premissas permitem. A partir de muitos casos observados, por exemplo, sobre a reprodução assexuada de planárias, infere-se que há um padrão ali delineado e pode-se concluir que há boas razões para alegar um enunciado geral segundo o qual todas as planárias se reproduzem assexuadamente. Todas as vezes que tentamos estabelecer verdades universais, ou quando procuramos dar uma explicação sobre as causas de determinados fenômenos, ou quando supomos ser capazes de prever o que ocorrerá com base no que a experiência registrou sobre casos em condições idênticas ou muito similares, estamos raciocinando indutivamente. As próprias leis científicas possuem tal universalidade e é por meio delas que os cientistas fazem previsões ou dão explicações unificadas para uma série de fenômenos (veremos um pouco mais sobre leis científicas na seção 2.1). Mas tais leis só podem ser enunciadas, se estiverem ancoradas em grande quantidade de evidência empírica que as endosse e na ausência de evidências em contrário. Ocorre, entretanto, que apenas de posse da forma lógica de um raciocínio indutivo, não é possível alegar que as premissas, quando verdadeiras, ofereçam mais do que algum grau de probabilidade de a conclusão ser verdadeira. Ao contrário da dedução, há elementos não formais que precisam estar disponíveis para dar suporte à conclusão que as premissas visam endossar. Isso pode ficar mais claro, ao compararmos dois raciocínios indutivos com forma idêntica, como aparecem no quadro abaixo:

Indução 1	Indução 2
<ul style="list-style-type: none"> ● Todos os metais observados até hoje dilataram-se sob a ação do calor. (premissa) ● Um grande e variado grupo de tais casos foi examinado, sem ocorrência de exceção. (premissa) ● Logo, provavelmente todos os metais se dilatarão sob a ação do calor (conclusão) 	<ul style="list-style-type: none"> ● Todos os cavalos observados até hoje nasceram antes de 2030 (premissa) ● Um grande e variado grupo de tais casos foi examinado, sem ocorrência de exceção. (premissa) ● Logo, provavelmente todos os cavalos nascerão antes de 2030. (conclusão)

Enquanto a indução 1 parece um bom modo de garantir a ideia de que metais se dilatam quando aquecidos (e continuarão a se comportar assim), a indução 2, que possui a mesma forma, não passa em nossos critérios intuitivos de um bom raciocínio. É muito provável que cavalos continuem a nascer depois de 2030, ainda que seja verdade hoje (escrevo em 2020) que todos os cavalos até agora tenham nascido antes de 2030. Descobrir quais seriam as condições necessárias para tornar uma conclusão obtida indutivamente mais provável (e em que grau) é parte importante da epistemologia da ciência.

O escocês David Hume (1711-1776) procurou analisar o tipo de raciocínio que conferia garantia para o conhecimento obtido mediante os sentidos, isto é, *a posteriori* (em oposição ao conhecimento *a priori*, que na visão do autor seria apenas de ordem lógica e, portanto, não dependente de qualquer dado sensorial). Hume percebeu que procedemos por indução sempre que observamos repetidos fenômenos com grande semelhança, seguidos de efeitos também semelhantes entre si, em conjunto com o hábito de esperar a continuidade dessa repetição em novas observações. A indução estaria por trás de qualquer tentativa de fazer inferências ampliativas, isto é, aquelas inferências em que ou supomos descobrir um fato generalizado, ou pretendemos descobrir causas para fenômenos a serem explicados, ou ainda quando queremos fazer uma previsão. Em todos esses casos, supõe-se que a natureza é regida por uma uniformidade, uma espécie de lei segundo a qual o universo se comporta sempre

da mesma maneira, de modo que o futuro será como o passado. Se a natureza for uniforme, uma série de observações sobre a queda de corpos, por exemplo, permite inferir, caso não haja exceção constatada, que todos os corpos caem quando abandonados próximos da superfície terrestre. Permite prever que o próximo corpo a ser abandonado nas mesmas condições comportar-se-á de modo similar. Permite supor que sem a causa, no caso a gravidade, não se seguiria o efeito, a queda livre. Mas para justificar a uniformidade da natureza seria necessário recorrer à indução: a natureza é uniforme porque até hoje ela se comportou dessa forma. Isso criaria um raciocínio circular segundo o qual a indução depende da uniformidade da natureza, mas a uniformidade da natureza também depende da indução. Uma vez que não temos uma justificativa independente para nenhum desses pressupostos, estaríamos sem condições de explicar que tipo de garantia a ciência dá para seus enunciados. Afinal, quantas observações seriam necessárias para confirmar uma teoria? E como garantir sem circularidade que uma teoria foi confirmada? Supomos que uma lei científica tenha a forma 'todo A é B', por exemplo: 'toda amostra de H₂O destilada entra em ebulição a 100°C no nível do mar'. Depois de muitos registros de 'As' que são 'Bs' (várias amostras de água fervendo a 100 graus no nível do mar) não podemos validamente inferir que é um caso geral, mas apenas que todos os casos até agora observados assim se comportaram. Dizer que sempre será assim é ultrapassar o que foi observado. Esse problema foi particularmente ressaltado pela crítica de Popper ao positivismo lógico e o retomaremos muito em breve, quando veremos as repostas do falsificacionismo e do bayesianismo.

1.2. Falsificacionismo: uma abordagem dedutivista da ciência

Popper, ciente do problema da indução e defendendo a indissociabilidade entre observação e teoria, acreditava que seria possível responder ao problema da indução e ao da demarcação com apenas um critério, chamado por ele de critério da falseabilidade ou refutabilidade. A ideia de que a ciência caracterizar-se-ia por observações teoricamente neutras coletadas por cientistas absolutamente imparciais, que chegariam às leis e enunciados gerais da ciência, operando por indução, não faria jus à lógica da investigação científica, nem à realidade dos cientistas. Popper está de acordo com positivistas de que o critério de demarcação tenha a ver com a testabilidade de um enunciado. Mas o

critério verificacionista não seria capaz de delimitar corretamente o terreno da ciência, distinguindo-a de criações teóricas pseudocientíficas. Popper estava preocupado com o fato de várias teorias infalíveis (algo que a ciência certamente não é) serem apresentadas como ciência, bastando, para isso, fazer ajustes *ad hoc* (isto é, com a finalidade de salvar a teoria) quando algum fato empírico parecer refutá-las. A ideia popperiana é considerar que um enunciado é testável se e somente se existir a possibilidade de que tal enunciado entre em conflito com observações possíveis ou concebíveis. Neste caso, o que caracterizaria a ciência é a *possibilidade* de que as hipóteses sejam testadas, ou seja, falseadas por testes empíricos.

É muito importante não confundir falseabilidade com falsidade. Não é objetivo da ciência fornecer um conjunto de enunciados falsos, isto é, enunciados que foram refutados por testes empíricos. Popper exigia algo diferente: que os proponentes de uma conjectura científica explicitassem as condições empíricas em que seria necessário abandonar aquela conjectura em favor de uma alternativa. Isso explica também a expressão “observações concebíveis”. Muitas propostas teóricas só poderiam ser refutadas com tecnologia não disponível atualmente, como é o caso da hipótese de que não existe nada mais veloz do que a luz, da teoria da relatividade. É, no entanto, possível imaginar um experimento que refute a conjectura relativista. O critério de Popper seria, portanto, um impulso para encontrar fenômenos que no lugar de confirmar uma hipótese, contribuiriam para sua ruína, caso ocorressem. É uma afirmação de que a ciência é uma prática crítica, disposta a ser revista sempre que isso se mostrar coerente com as evidências empíricas refutadoras. As pseudociências, os credos religiosos e políticos, e as teorias metafísicas em geral não apresentam esse caráter, sendo, ao contrário, dogmáticos, lançando mão de mecanismos retóricos, adendos e novas desculpas que acomodariam os fatos refutadores. A astrologia, para dar um exemplo batido de pseudociência, faz previsões genéricas o bastante para caber na vida de qualquer pessoa, sendo ela ou não daquele signo regente. Ainda assim, algumas previsões astrológicas podem falhar. Nesses casos de falha de previsão, no lugar de sugerir o abandono da astrologia, os astrólogos tentarão mostrar que as falhas eram explicadas pelo signo ascendente; em caso de falhas também relacionadas ao ascendente a questão seria atribuída a outro astro e assim por diante. Recuando em cada nova explicação, chegamos, por fim, num sistema em que a

personalidade suposta para cada signo está potencialmente presente em todo e qualquer indivíduo, seja de qual signo for. Vista dessa forma, a astrologia é irrefutável. Uma observação: o conceito de pseudociência se aplica somente quando os praticantes insistem que suas práticas são científicas, o que talvez não seja o caso da astrologia contemporânea. Mas o exemplo citado é facilmente adaptável à medicina alternativa e a certas psicoterapias. Sobre o criacionismo e o projeto inteligente como rivais da teoria evolutiva, a dificuldade é a inexistência de qualquer fato concebível que faça os adeptos reconhecerem que sua teoria de estimação é falsa. As ciências, de acordo com Popper, podem ser refutadas empiricamente e seu método seria mais ou menos similar ao que se segue:

i	descoberta de um problema científico;
ii	elaboração de uma conjectura que acoberte o problema inicial;
iii	submissão da conjectura a testes de falsificação;
iv	se a conjectura não foi falsificada, é considerada corroborada (por enquanto) e os passos (iii) e (iv) continuam a ser realizados;
v	se houve falsificação, isso transforma-se num problema científico e todos os passos são novamente realizados com a nova conjectura decorrente desse processo.

A nova conjectura precisa explicar todos os fatos que a conjectura anterior já explicava, além de cobrir o fato refutador da antecessora, bem como prever fatos adicionais, que, não ocorrendo, possibilitariam nova refutação. Não por acaso, uma das obras mais famosas de Popper sobre o método científico teve o título de *Conjecturas e refutações*.

De um ponto de vista lógico, o que Popper propôs nada mais era do que a aplicação, no domínio das ciências experimentais, de um tipo de raciocínio dedutivo conhecido pelo nome latino de *modus tollens*, um velho conhecido de lógicos e matemáticos. O *modus tollens* tem como premissas uma condicional (Se *A*, então *B*) e a negação do conseqüente da condicional (não-*B*). A conclusão dedutiva é a negação do antecedente (não-*A*).

Se A então B	(premissa)
Não- B	(premissa)
Logo, não- A	(conclusão)

Sendo A uma conjectura ou hipótese e B uma consequência esperada dessa hipótese, caso a mesma seja verdadeira, a não ocorrência de B impõe a conclusão de que A é falsa. Por outro lado, se a consequência ocorrer, isso não permite qualquer conclusão lógica sobre A , dado que uma hipótese falsa pode ter consequências verdadeiras. Um exemplo trivial deste segundo caso: se o professor Tiago possui alto grau de hipermetropia, então ele precisa de óculos. De fato, o professor Tiago precisa de óculos, mas não se pode disso concluir que o professor Tiago tenha hipermetropia. Seus óculos podem, por exemplo, ser para miopia ou presbiopia, ou ainda para corrigir uma sensibilidade à luz. Obviamente, constatar que o professor Tiago não precisa de qualquer tipo de lente corretiva é indício conclusivo de que não possui hipermetropia.

O que Popper sugere, portanto, é que a lógica subjacente à pesquisa científica não seria, de fato, indutiva, mas uma dedução análoga ao *modus tollens*, em que uma refutação dá uma resposta precisa sobre a falsidade da conjectura testada e uma confirmação conta apenas como mais um tempo de sobrevivência àquela conjectura. As conjecturas que vão resistindo às tentativas de falseamento são consideradas corroboradas, mas nunca podem ser admitidas como definitivamente verdadeiras, já que não há testes que poderiam garantir isso. Dessa forma os enunciados universais, tais como as leis científicas (alguns exemplos são as leis de Newton, a lei de Kepler, a lei de Boyle, as leis de Mendel) possuem esse status devido ao fato de terem resistido a um número tão grande de testes que já ficou patente seu valor. Ainda assim, não há garantias de que tais leis sejam definitivas e, a bem da verdade, algumas das mencionadas já cederam seu lugar para novas leis e teorias mais abrangentes (é o caso da relatividade no lugar da gravitação de Newton e da teoria cinética dos gases em relação a lei de Boyle). Leis são enunciados universais que possuem a forma 'todo A é B '. Enunciados desse tipo seriam refutados por ocorrências singulares, de um A que não seja B . O grau de testabilidade de um enunciado é proporcional a quanto ele proíbe. Por essa razão verdades lógicas (tautologias) sequer teriam base empírica, já que nada no mundo as tornaria falsas, ao passo que enunciados particulares como 'algum A é B ' só seriam refutados pela inexistência

de *As* que sejam *Bs*, o que seria praticamente impossível de provar (muitos *As* que não são *Bs* não o fazem, a menos que se queira voltar ao problema da indução).

Assim como o verificacionismo, também a proposta de Popper apresenta certos problemas, razão pela qual alguns filósofos da ciência consideram o falsificacionismo uma descrição inadequada do método científico. Dentre as várias críticas, três parecem mais severas.

A primeira crítica é a de que Popper não descreve como a ciência é, mas como deveria ser. Os filósofos retratam esse problema como a distinção entre o nível prescritivo/normativo e o nível descritivo/factual. Em algum momento o defensor do falsificacionismo precisaria escolher se o método de conjecturas e refutações é o que os cientistas fazem ou se é um ideal que os cientistas deveriam seguir, embora não o façam sempre. Em vários episódios da história da ciência, o que se observou foi um apego à teoria refutada no lugar de seu abandono. A órbita de Urano refutava as previsões da lei da gravitação universal, de Newton. Entretanto, no lugar de abandonar a teoria, os cientistas criaram uma hipótese *ad hoc*, de que deveria haver outro planeta desconhecido, o que explicaria a anomalia. Imaginando que não era possível uma teoria com tão grande sucesso ser falsa, os astrônomos preferiram se apegar à mesma e calcularam a posição do suposto astro desconhecido, chegando à descoberta de um novo planeta do sistema solar, Netuno. E mesmo que a descoberta de Netuno faça pesquisadores recusarem que tenha havido refutação da gravitação de Newton, a órbita de Mercúrio insiste em não corresponder à teoria do físico inglês, sem que até hoje se saiba o porquê. É controverso se esse capítulo da astronomia conta mesmo como uma falsificação, mas ele revela que uma falsificação pode não ser tão simples como parece e que muitos cientistas estão tão comprometidos com suas conjecturas que não abririam mão delas facilmente. Seria prudente acusar tais pesquisadores de não fazer ciência? Sustentar que a falseabilidade é o critério de demarcação pode vir a deixar de fora muitos casos que hoje consideramos de sucesso científico. Desse modo, talvez o critério não sirva para demarcar ciência em relação a outros discursos, mas para sugerir quais são as práticas recomendadas para se fazer boa ciência (na medida que o dogmatismo, o apelo irracional a uma teoria ou hipótese e o viés da confirmação podem nos fazer defender propostas teóricas ruins e rejeitar alternativas melhores). Isso, argumenta-se, não é fazer epistemologia, mas axiologia da

ciência, uma vez que o que está em jogo são os valores que tornam uma ciência bem (ou mal) conduzida. Esteja isso certo ou não, o fato é que tanto a boa ciência quanto a má são, no fim, instâncias da ciência e o critério de demarcação por falseamento não parece adequado.

Outra crítica ao falsificacionismo é a de que, ao abrir mão do papel da indução, não se consegue explicar o sucesso do empreendimento científico em fazer previsões acuradas. Se dissermos que as teorias científicas são apenas conjecturas que sequer podemos supor serem provavelmente verdadeiras, deixamos de lado o aspecto pragmático que mais chama atenção da sociedade no que toca a respeitabilidade e confiabilidade do conhecimento científico. Crenças (entendidas como opiniões) são não apenas elaborações mentais, mas também guias para a ação. Decisões no nível pessoal e social são tomadas tendo em mente o que se considera verdadeiro ou muito provável. E não teríamos a menor ideia do que esperar de nossas teorias e dos artefatos tecnológicos nelas inspirados, se não tivéssemos ao menos alguma segurança de que as coisas sairiam muito próximas das previsões científicas. Popper nos dá uma regra para estabelecer quais conteúdos científicos são falsos, mas não permite afirmar nada sobre a solidez do atual estágio do conhecimento. Como seria possível explicar a erradicação de várias doenças depois da administração generalizada de vacinas por uma geração, ou como fenômenos sequer imaginados antes poderiam ser previstos e confirmados sem que as teorias que permitiram tais previsões não fossem consideradas ao menos, muito próximas de serem verdadeiras? Veja, por exemplo, um caso absolutamente importante na história da física recente. Em 1916, Einstein publicou uma teoria que incorporava a gravidade em sua relatividade especial. O modelo, chamado então de relatividade geral, previa que grandes massas causassem distorção no espaço-tempo, de modo que mesmo a luz seria afetada pela curvatura espaço-temporal quando próxima de corpos muito massivos, fazendo, assim, com que os fótons tivessem sua trajetória alterada. Essa predição contrariava boa parte do que se conhecia sobre a natureza da luz e testá-la era muito difícil por causa de dificuldades técnicas, como o fato de a luz das estrelas ser ofuscada pelo sol, sendo este o corpo massivo a criar a distorção. Seria, entretanto, possível confirmar as previsões de Einstein num eclipse solar total. O experimento foi realizado no Brasil, na cidade de Sobral, no Ceará, em 1919. As fotografias da luz das estrelas tiradas durante o eclipse, comparadas com as observações de

quando as mesmas estrelas ocupavam o mesmo lugar no céu sem, entretanto, a interferência do sol, confirmaram as suspeitas de Einstein. Já é surpreendente que uma teoria que pode vir a mostrar-se falsa permita que calculemos correta e precisamente a data e o local de onde seria possível observar um eclipse total. Embora possa nos espantar, não é nenhuma novidade a previsão de eclipses. Conta-se que Tales de Mileto foi capaz de prever um, há cerca de 2600 anos e, ao que parece, civilizações egípcias, babilônicas e pré-colombianas também o faziam. Mais surpreendente ainda é que um fenômeno completamente contraintuitivo venha a ser observado depois de ser previsto por uma nova teoria. Se tomarmos o que Popper sugere, isso conta apenas como uma trivialidade de que tal teoria não foi provada falsa, ainda. Para estabelecer causas, criar modelos explicativos e preditivos, para estabelecer leis científicas gerais, não parece suficiente manter que teorias científicas são apenas conjecturas corroboradas. Cientistas estão bastante convencidos de que teorias capazes de sucesso surpreendente são modelos aproximados da realidade e isso requer alguma confiança em raciocínios indutivos.

Uma última objeção ao falsificacionismo é de ordem lógica e remonta a dois autores que, cada qual ao seu modo, procuraram minar a ideia de que seja possível testar e modificar hipóteses isoladas. Este problema ficou conhecido como problema de Duhem-Quine. Pierre Duhem (1861-1916) foi um físico e filósofo francês e Willard Quine (1908-2000) foi um lógico estadunidense. Sem entrar em minúcias, a ideia que leva ao problema de Duhem-Quine é a de que uma falsificação coloca em xeque todo um conjunto de crenças que atuavam em interdependência, tal como numa rede com vários nós, cada qual ligado a outros nós. Já mencionamos neste texto, quando tratamos de uma dificuldade do positivismo lógico, sobre como cada hipótese científica é dependente de outras teorias científicas e demais crenças de fundo. Quando a consequência prevista é derivada de um conjunto de crenças, o máximo que podemos afirmar é que ao menos uma das crenças de tal conjunção é falsa. Qual delas? Não é possível dizer. O *modus tollens* que estaria por trás da investigação científica seria um pouco mais complexo do que imaginara Popper. O raciocínio tomaria, de acordo com a tese de Duhem-Quine, uma forma similar à seguinte, onde H é hipótese central, A são hipóteses auxiliares, C são as crenças de fundo e P a previsão a ser refutada:

Se ($H \& A \& C$), então P	(premissa)
Não- P	(premissa)
Logo, não é o caso que ($H \& C \& A$)	(conclusão)

O conjunto formado por $H + A + C$ é falso tanto se apenas H for falsa, quanto se apenas A for falsa, ou ainda se a proposição falsa for C , ou se mais de um dos elementos for falso simultaneamente. Em último caso, todos eles podem ser falsos. Desse modo, é tão racional abandonar a hipótese em questão, mantendo todos os outros nós da rede intactos, quanto modificar as demais crenças de fundo, conservando a hipótese central.

1.3 Abordagem do pluralismo historicista

Alguns filósofos da ciência, bastante apegados à história de como as ideias científicas nasceram e se desenvolveram, procuraram apresentar uma alternativa a positivistas e falsificacionistas quanto ao entendimento do que seria a natureza da ciência. Antes, porém, de tratar dos autores historicistas, é importante situar o leitor sobre dois pontos de convergência entre verificacionismo e falsificacionismo que serão atacados por aqueles autores: ambas as epistemologias da ciência até aqui descritas se apoiam na distinção entre contexto de descoberta e contexto de justificação e ambas sustentam que só há um método científico (embora discordem sobre qual é esse método).

Embora o verificacionismo e o falsificacionismo discordem sobre a importância do raciocínio indutivo, em ambas as abordagens está presente uma distinção entre o modo como os cientistas descobrem ou inventam suas teorias e o modo como eles justificam tais teorias à luz dos fatos empíricos. Essa é uma distinção entre dois contextos: o da descoberta e da justificação. A ideia associada à objetividade e pertinência das hipóteses teria pouco ou nada a ver com as condições psicológicas, históricas, sociais etc. dos cientistas que as formularam e sim com o mérito das próprias hipóteses em explicar, prever, unificar e permitir avanços tecnológicos. Poderíamos fazer uma analogia com uma obra de arte: o processo criativo do artista, embora possa interessar à história da arte e à psicologia da criação, não é o que faz uma obra ser genial e sim a forma, a intensidade, a complexidade e a beleza ali retratadas. Se perguntarmos a um crítico musical os motivos pelos quais julga uma melodia

genial, provavelmente a explicação repousaria nos últimos aspectos e não no fato de o autor criar de madrugada ou de tarde, de ficar dias sem banho ou recolher-se em meditação por um ano inteiro. Um dos herdeiros do positivismo lógico, o alemão Carl Hempel (1905-1997) sugeriu, em seu livro *A filosofia da ciência natural* (publicado originalmente em 1966), que hipóteses podem ser livremente inventadas, mas nunca livremente aceitas. Isto é, a parte propriamente indutiva do método científico não seria aplicável ao contexto da descoberta. Neste último, o cientista estaria desimpedido para criar, e o faz seguindo critérios que muitas vezes sequer podem ser considerados racionais. Mas uma vez formulada a hipótese, a mesma terá uma série de consequências lógicas que podem, por sua vez, ser verificadas (ou eventualmente refutadas) empiricamente. Seriam, na opinião de Hempel, essas consequências empiricamente verificadas que dariam suporte indutivo à crença de que a hipótese é verdadeira. Para Popper, a situação é similar, com a ressalva de que os testes são feitos para refutar a hipótese e que as corroborações não seriam suficientes para atestar a veracidade da hipótese. O próprio termo ‘conjectura’ escolhido por Popper dá a entender um processo criativo, não sujeito a regras lógico-metodológicas. Tudo que importa é sujeitar as hipóteses ao escrutínio da experiência.

Um outro ponto de convergência entre o positivismo lógico e o falsificacionismo é a ideia de que haveria apenas um método científico capaz de delimitar claramente o domínio da ciência, diferenciando de outros domínios de crença. A abordagem que vamos examinar a seguir questiona tanto essa ideia de monismo metodológico quanto a pertinência da separação entre contexto de descoberta e contexto de justificação. Na visão dos autores pluralistas/historicistas, seria inadequado discursar sobre o que é ciência sem prestar atenção aos vários modos como essa atividade se apresentou ao longo da história. Supor um método único e escolher cuidadosamente exemplos históricos que endossem essa escolha, além de deixar de fora uma série de procedimentos e teorias que poderiam ser, pelo menos no contexto de uma época, considerados legitimamente científicos, também é um mal-uso da história da ciência, uma farsa.

Três obras foram absolutamente importantes para quebrar o monopólio de falsificacionistas e verificacionistas na disputa para responder à questão sobre o que é ciência: *Patterns of Discovery* [Padrões da descoberta], publicado

em 1958 por Norwood Hanson (1924-1967), *A estrutura das revoluções científicas*, publicado em 1962 por Thomas Kuhn e *Contra o método*, de Paul Feyerabend, cuja publicação se deu em 1970. O primeiro livro procurou identificar, com a ajuda da psicologia gestaltiana, uma lógica da descoberta científica, mostrando como as observações, longe de serem teoricamente neutras, estavam carregadas de pressupostos teóricos, motivadas pelo conhecimento prévio e expectativas inconscientes, além da obsessão por encontrar padrões na natureza, o que explicava a preferência pelas explicações mais simples e unificadoras. Alguns exemplos retirados da *gestalt* (palavra alemã que significa 'forma') ilustram o argumento de Hanson de que observar algo é vê-lo como isto ou aquilo:

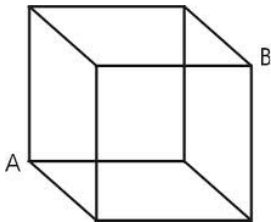
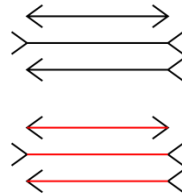


Figura 1 - Cubo de Necker



2 - Ilusão de Muller-Lyer

A figura 1 retrata um cubo cuja face mais próxima do observador pode ser percebida inclinada para cima (A) ou para baixo (B). O interessante sobre essa figura é que as pessoas que a enxergam com uma determinada inclinação e as pessoas que a enxergam com a outra inclinação estão vendo exatamente os mesmos traços. Já a figura 2 mostra como a percepção dos traços paralelos pode ser diferente a depender de como as setas laterais são posicionadas. Os três traços da figura 2 possuem o mesmo comprimento, mas nossa observação não consegue identificar tal equivalência por causa de nossa tendência psicológica de ver as imagens como um todo, impingindo pressupostos, completando a figura. Tais ilusões mostram a dependência teórica mesmo de observações simples. Mas isso fica bastante evidente quando se comparam as observações feitas por um leigo com as de um especialista utilizando o mesmo método de detecção. Um exame de ultrassom, por exemplo, pode revelar a existência de cistos no ovário, ou um embrião ainda em fase inicial de formação,

caso analisada por um médico especializado. A mesma imagem, entretanto, vista por uma pessoa sem o treinamento adequado não parece mostrar mais do que manchas e rabiscos indistintos. É o treino e o conhecimento teórico prévio que permite ao especialista ver o que ninguém mais vê, se não tiver a mesma preparação. Esse parece ser o caso de inúmeros procedimentos de observação realizados no âmbito de uma ciência especializada.

Kuhn e Feyerabend assumiram a tese hansoniana da dependência teórica da observação e procuraram mostrar que as escolhas por uma teoria científica, bem como do método pertinente a cada período da ciência poderiam variar bastante, conforme o contexto. Mais ainda, eles pensavam que em muitos casos, o próprio modo de compreender a realidade sofreria uma reviravolta tão grande nos esquemas mentais que guiavam os cientistas, que sequer poderia haver algum tipo de comparação objetiva entre duas teorias científicas que não compartilhassem dos mesmos pressupostos. Kuhn entendia que a ciência em seu período normal e estável (em contraposição aos momentos de revolução científica) era uma prática de solucionar quebra-cabeças desenvolvida a partir de um núcleo de consensos que guiava a comunidade científica. Tal consenso era motivado pelo despontar de uma nova ciência, cujo sucesso motivava os demais pesquisadores a espelhar o mesmo método, e a debruçar-se sobre os problemas indicados, mas ainda não respondidos pela ciência bem-sucedida. Essa concordância sobre o método e os problemas relevantes da ciência, característica da ciência normal, foi chamada por Kuhn de “paradigma”. Depois que um tipo de investigação se torna paradigmático, a comunidade científica pode ampliar os problemas e trabalhar no quebra-cabeças ali delineado, julgando o grau de sucesso das tentativas, à luz do paradigma que compartilham. Isso seria análogo a um jogo. Na visão de Kuhn, cientistas não testam paradigmas, isto é, as regras segundo as quais resolvem o quebra-cabeça científico, mas apenas jogam conforme tais regras paradigmáticas permitem. O enxadrista não estaria interessado em criar um jogo diferente, mas em conseguir derrubar o rei adversário, movendo as peças do modo como a regra estabelece. Obviamente, dentro de um mesmo paradigma, é possível falar de crescimento do conhecimento científico, de acúmulo de problemas solucionados, de progresso. Mas o mesmo não pode ser dito em relação a revoluções científicas, nas quais a comunidade de pesquisadores deixa de perceber os problemas de um paradigma obsoleto como relevantes, rejeitam os métodos antigos e voltam-se

para novos problemas não disponíveis antes do surgimento do novo paradigma. Aí sim há uma troca do jogo a ser jogado. Revoluções científicas seriam períodos turbulentos e breves nos quais alguns paradigmas competem entre si até que um deles vença a competição, estabelecendo as bases para a nova forma de fazer ciência. Para que seja possível uma mudança paradigmática, é necessário que o paradigma vigente acumule anomalias, problemas que os métodos e as linhas de investigação em vigor não sejam capazes de acomodar, explicar, eliminar ou responder. A desconfiança com o paradigma vigente pode aumentar à medida que alternativas teóricas pareçam dar conta daquelas anomalias. Os problemas do paradigma abandonado e sua linguagem seriam, em sua maioria, esquecidos, considerados não relevantes e mesmo incompreensíveis. Apenas alguns termos e problemas, na medida em que fossem considerados precursores do novo paradigma, seriam mantidos. Ainda assim, os próprios significados daqueles termos e problemas seriam alterados substantivamente para parecer que a ciência do paradigma atual é um acúmulo progressivo do que havia de verdadeiro nos paradigmas do passado mais remoto. Kuhn considerava que paradigmas sucessivos seriam como duas línguas tão diferentes que a tradução seria demasiadamente imperfeita. A linguagem de um paradigma remeteria a uma realidade incapaz de ser captada pela linguagem de outro paradigma. É como se ambas se referissem a mundos completamente distintos. Revoluções científicas foram as que ocorreram, por exemplo, na passagem do modelo geocêntrico para o heliocêntrico na astronomia, da química do flogisto (ainda muito dependente de conceitos da alquimia) para o modelo de Lavoisier (com a criação da tabela periódica como conhecemos) e quando vários modelos biológicos são substituídos pela teoria da evolução, criando o modo atual de fazer biologia. A compreensão dos modelos abandonados para os cientistas atuais é obscurecida pela diferença gritante entre eles. Pense, por exemplo, na palavra que nomeia nosso planeta: 'Terra'. A ideia de que a Terra está em movimento contraria o próprio significado do nome, que estava associado à fixidez, antes de Copérnico. Não é sem razão que Copérnico foi rechaçado pelos adeptos do modelo cosmológico anterior, pois, para os últimos, seria como afirmar que o círculo deveria ser entendido como quadrado. É como se a visão alternativa que Copérnico propôs alterasse a própria compreensão da realidade, em um processo análogo ao que ocorre quando passamos a ver o cubo de Necker de modo diferente. Outro exemplo

disso são as expressões 'nascer do sol' e 'pôr do sol', que precisaram ser reinterpretadas metaforicamente, uma vez que é o planeta que está girando e não o sol se movendo. Também o termo 'massa', considerado uma propriedade dos corpos na mecânica newtoniana, foi reinterpretado na mecânica relativística como uma relação com a energia. Essa incapacidade de tradução recebeu de Kuhn o nome de incomensurabilidade. Feyerabend também utilizou o conceito de incomensurabilidade para tratar de termos de teorias passadas que não poderiam ser incorporados em teorias científicas mais gerais e mais atuais. O conceito de *impetus*, responsável por explicar movimentos de projéteis na ciência medieval de inspiração aristotélica, não possui um equivalente termo newtoniano, já que na nova visão, o esperado é que se nenhuma força estiver atuando, um corpo em movimento permaneceria em movimento com velocidade constante (lei da inércia). Assim como Kuhn, Feyerabend também utilizou a história da ciência para mostrar como o contexto de descoberta (com todos os seus elementos retóricos, crenças de fundo, motivações não epistêmicas, valores etc.) era tão importante para o estabelecimento de mudanças teóricas relevantes quanto os elementos de justificação. Além do mais, ficaria evidente que a ciência não poderia seguir um método único, já que nenhuma teoria científica poderia concordar com todos os fatos, e já que os fatos não poderiam ser estabelecidos sem qualquer tipo de carga teórica. Na visão de Feyerabend, muitas teorias persistiram depois de falsificadas, outras ainda teriam sido substituídas sem qualquer fato refutador, mas apenas pela existência de uma teoria alternativa. A refutação muitas vezes só viria depois de uma teoria já ter sido há muito descartada. Impedir a nova teoria de se desenvolver tornaria impossível, em alguns casos, sequer descobrir fenômenos potencialmente refutadores de teorias anteriores. Se fosse necessário, portanto, derivar uma norma metodológica que se encaixe em todos os relatos da ciência, em toda e qualquer época da ciência, tal norma seria "tudo vale" (o que explica o título de sua famosa obra). Para Feyerabend, o melhor modo de potencializar o crescimento do conhecimento científico seria multiplicar métodos e alternativas teóricas, permitindo que as mesmas alcancem maturidade no lugar serem descartadas por um processo de falsificação prematuro ou por inconsistências com teorias já muito bem-aceitas. A epistemologia historicista da ciência, através da ideia de incomensurabilidade de paradigmas, e da observação de que haveria muitas práticas que em algum momento foram tidas por científicas, embora

distintas das nossas concepções atuais, faz com que seja necessário pensar pluralisticamente o conceito de ciência. Sob muitos aspectos, esses autores também podem ser considerados relativistas, pois o que é ciência dependeria da época, do paradigma, da relevância social e de vários elementos não passíveis de serem expressos numa lógica de demarcação e mesmo de racionalidade. Recusar as advertências dos filósofos historicistas da ciência é deixar de fora muitas construções teóricas que constituíram, no passado, o que havia de melhor cientificamente. E aquilo que hoje considera-se um erro ou uma superstição do passado era sustentado e elevado ao status de ciência legítima pelos mesmos processos sociais que fazem com que a ciência atual seja considerada legítima, embora a ciência do passado e a atual sejam muito diferentes na linguagem, no método e no conteúdo. Assumir a ideia de incomensurabilidade de paradigmas tem como consequência não só a pluralidade de métodos que podem ser reconhecidos na história de cada ciência particular, mas também o fato de que tal história não é cumulativa, com a consequência de tornar problemático o conceito de progresso entre paradigmas. Neste caso, também não seria possível um critério puramente racional que justifique a predominância de um paradigma atualmente. Se um estudioso atento à história da ciência, portanto, quisesse oferecer uma definição de ciência, ele deveria optar por considerar que a atividade não possui uma única característica diferenciadora. Ele poderia, neste caso, considerar aplicar o conceito de semelhança familiar, proposto por Wittgenstein para atividades como jogos e religiões. Tais conceitos reuniriam uma família de elementos que seriam parecidos uns com os outros em aspectos diferentes, de modo que não seja possível uma demarcação rígida, isto é, a apresentação de uma condição necessária. É como uma família humana, em que um dos filhos se parece com a avó, o outro com o tio paterno e a caçula com a mãe, sem que todos se pareçam com todos em um aspecto específico. Assim funcionaria, segundo Wittgenstein, o modo como classificamos os jogos (e as religiões). Há algo parecido entre um jogo de cartas e o dominó; outra coisa liga o dominó a um jogo de dados, os dados a um jogo de tabuleiro, este ao xadrez etc. O mesmo se daria com a atividade científica, que se assemelharia com várias práticas contemporaneamente superadas, mas consideradas científicas no passado, tais como a filosofia pré-socrática, os estudos aristotélicos sobre o movimento e sobre os seres vivos, os estudos de alquimia etc.

Problemas na abordagem historicista também abundam. Provavelmente o leitor já deve imaginar que muitos cientistas e filósofos da ciência não estejam dispostos a assumir que a atividade científica seja irracional, embora muitos viessem a reconhecer que suas explicações sobre a natureza da ciência precisavam incorporar a história da ciência e acomodá-la de modo consistente em suas teorias. Uma das críticas dirigidas à ideia de incomensurabilidade de paradigmas é a de que se o paradigma emergente é capaz de solucionar anomalias do anterior, então ambos não são, de todo, incomensuráveis. Neste caso o termo 'incomensurável' dá a entender algo muito mais exagerado do que a explicação historicista de Kuhn permite inferir.

Outro ponto que é bastante problemático na tentativa de solução (ou talvez de dissolução) do problema da demarcação por meio do historicismo é o perigo de auto-refutação. A história, enquanto ciência, também não seria regida por paradigmas, segundo os quais determina-se como os fatos do passado devam ser estudados e como devam ser reconstruídos? Colocar a história da ciência em melhor posição para abordar o problema da demarcação do que as reconstruções racionais da filosofia e dos próprios cientistas na compreensão de sua atividade parece conferir à história (enquanto ciência) uma objetividade que ela nega às ciências naturais que toma por objeto. Se Kuhn estiver certo, sua própria abordagem não possui nenhuma vantagem objetiva em propor que a história da ciência seja não cumulativa, afinal ele fala a partir de um paradigma, ao passo que uma visão de progresso cumulativo fala de outro paradigma. Se os paradigmas não puderem ser objetivamente comparados, a tese cumulativista não é objetivamente pior que a incomensurabilista. Como alternativa, Kuhn precisaria defender que a incomensurabilidade é uma verdade objetiva, ou seja, um dado da realidade capturado pelo historiador da ciência e, neste caso, esse dado não seria relativo a um paradigma. Mas se a história for capaz de captar dados objetivos e independentes de paradigmas, por que não as ciências naturais?

Uma última palavra sobre a questão da demarcação e sobre as abordagens historicistas que merece ser mencionada é o fato de haver tentativas conciliatórias de uma lógica da demarcação, com as intuições advogadas pelos filósofos historicistas da ciência. Imre Lakatos (1922-1974) sugeriu que a racionalidade de um programa de pesquisa não está ligada ao progresso, ou seja, ao acréscimo de conteúdo empírico. Este último só pode ser reconhecido

retroativamente. Os cientistas seriam falsificacionistas (e, portanto, racionais), mas tal procedimento não equivale ao ingênuo e simplório *modus tollens* sugerido na explicação sobre Popper. Num programa de pesquisa haveria um núcleo duro que os pesquisadores ligados ao programa não estariam dispostos a negociar, enquanto dirigiriam as refutações e reformulações ao cinturão de hipóteses auxiliares e crenças de fundo que envolvem o núcleo. Os programas de pesquisa seriam progressivos, se se mostraram fecundos com o tempo, e regressivos, se resultaram num beco sem saída. Mas não seria possível julgar qual programa de pesquisa é superior ao outro antes que a história assim o julgue. Larry Laudan (1941-) também procurou oferecer uma alternativa mediadora, considerando critérios objetivos de progresso em conformidade com a história da ciência. Para este último, entretanto, a questão não seria estabelecer o significado de racionalidade e depois identificar onde, na história, a escolha foi mais racional. Sua proposta foi reconhecer objetivamente o progresso como capacidade de resolução de problemas e derivar de casos incontestáveis de avanço na história da ciência uma noção de racionalidade. Para Laudan, portanto, não é o progresso que depende da racionalidade, mas a racionalidade que depende do progresso reconhecido.

1.4 Bayesianismo: uma solução por meio da probabilidade?

Voltemos agora ao problema da indução. Já adiantamos muitas coisas sobre essa questão de modo que é possível ser mais breve neste tópico. Dissemos que a indução consiste num tipo de raciocínio ampliativo, que nos permite ir além dos casos observados para formar generalizações, fazer previsões e dar explicações. Também já fizemos notar que boa parte do sucesso da ciência está na capacidade de oferecer as causas corretas dos fenômenos, isto é, prever com boa margem de precisão o que ocorrerá se determinadas condições naturais e/ou artificiais forem cumpridas, bem como o modo de atuação de uma substância no organismo, o modo como a eletricidade deve atuar para fazer um aparelho funcionar, os meios mais eficazes para transmissão de dados, os genes responsáveis por determinada característica etc. Todas essas exhibições do poderio científico são estabelecidas devido à habilidosa capacidade da ciência em fazer induções muito confiáveis. E, como também já vimos, não é possível oferecer uma justificação não indutiva para a indução.

Essa foi precisamente a razão para Popper ter varrido o problema para debaixo do tapete, sugerindo que a lógica sustentadora da prática científica era dedutiva. Nas deduções válidas, as premissas (os pontos de partida do raciocínio) garantem a conclusão, se forem verdadeiras. Isso faz com que, uma vez estabelecidas as premissas de uma estrutura válida de pensamento, cheguemos a um veredito conclusivo. O mesmo não pode ser dito das induções que, mesmo partindo de verdades dadas, no máximo, suporte não definitivo para aquilo que se visa defender. A ocorrência de um grande número de metais que conduzem eletricidade dá suporte não definitivo à conclusão de que todos os metais conduzem eletricidade, assim como as várias vezes que a água entrou em ebulição a 100°C dão suporte não definitivo para pensar que essa temperatura é um padrão. Uma ocorrência diferente pode modificar a força desse suporte à conclusão.

Há, entretanto, uma tentativa de justificação alternativa para a indução que tem sido bastante defendida atualmente. Trata-se do recurso à estatística e à probabilidade, aplicando o teorema de Bayes, algo que ficou conhecido como bayesianismo. Um pressuposto dessa nova valorização da indução é a de que o conhecimento empírico é suportado com graus diferentes. A dedução não admite graus, já que ou a conclusão é totalmente garantida pelas premissas ou não é. Mas o tipo de suporte que possui uma crença baseada na experiência varia conforme a quantidade e a qualidade da evidência disponível. Como a ciência empírica é uma atividade em constante desenvolvimento, o crescimento de boa evidência a favor de uma conclusão e/ou o crescimento de boa evidência contra uma conclusão faz variar respectivamente para mais ou para menos o grau de probabilidade de uma crença científica. Neste caso, portanto, a atitude científica deve acompanhar as mudanças operadas na probabilidade da crença pela evidência atualmente disponível e admitir revisão no grau com que uma crença deve ser suportada em acordo com as novas evidências que venham a surgir. O teorema de Bayes não é propriamente novo, o matemático britânico Thomas Bayes (1702-1761) teve seu artigo publicado dois anos após sua morte. O que é mais recente é a aplicação do teorema na justificação do conhecimento científico. Vejamos como, especificamente, pode-se fazer isso.

Os axiomas da probabilidade são bastante simples:

- ❖ A probabilidade de uma proposição A [ou $\text{prob}(A)$] é uma medida entre 0 e 1.
- ❖ Se A for uma verdade necessária, $\text{prob}(A) = 1$
- ❖ Se A for impossível, $\text{prob}(A) = 0$
- ❖ Se A e B forem mutuamente excludentes, $\text{prob}(A \text{ ou } B) = \text{prob}(A) + \text{prob}(B)$

Para exemplificar o último axioma, as chances de um dado lançado resultar nos números 1 ou 2 é de $1/6 + 1/6$ (0,333), isso porque resultar 1 exclui resultar 2 e vice-versa. Já as chances de duas rolagens independentes resultarem, cada uma, em um 2 é de $1/36$ (0,028), pois a $\text{prob}(A \text{ e } B) = \text{prob}(A) \times \text{prob}(B)$.

Em casos nos quais os eventos não são independentes, entra em cena a noção de probabilidade condicional, segundo a qual, no lugar de multiplicar a probabilidade independente de que B ocorra, multiplicamos apenas a probabilidade de B dado o fato de A ocorrer, isto é, multiplicamos pela $\text{prob}(B|A)$. Assim,

$$\text{prob}(A \text{ e } B) = \text{prob}(A) \times \text{prob}(B|A).$$

Este último caso é ilustrado pelo cálculo das chances de um lance de um dado que resulte em um número par e primo. São primos os números 1, 2, 3 e 5. Os pares (2, 4, 6) resultam na $\text{Prob}(\text{Par}) = 3/6 = 0,5$. E a probabilidade de um número ser primo dado o fato de ser par, $\text{prob}(\text{Primo}|\text{Par})$, é $1/3$ (0,333). (Dentre os três números pares, apenas um é primo, a saber, o número 2).

$$\text{prob}(\text{Par e Primo}) = \text{prob}(\text{Par}) \times \text{Prob}(\text{Primo}|\text{Par}) = 0,5 \times 0,333 = 0,166$$

Repare que o mesmo número seria obtido se calculássemos $\text{prob}(A \text{ e } B) = \text{prob}(B) \times \text{prob}(A|B)$. Neste caso, teríamos a chance de $4/6$ de rolar um número primo $\times 1/4$ de obter um par dado ser primo (só um par entre os 4 primos disponíveis). O resultado seria $1/6$ ou 0,166.

A probabilidade de A ou B quando A e B não são independentes é assim definida:

$$\text{prob}(A \text{ ou } B) = \text{prob}(A) + \text{prob}(B) - \text{prob}(A \text{ e } B)$$

A fórmula acima é mais ou menos evidente porque se não subtrairmos a

probabilidade conjunta de A e B, estaríamos computando essas chances duas vezes (a probabilidade dos As que são Bs já foram computadas na probabilidade de A).

Voltemos agora à probabilidade de A e B quando são eventos relacionados:

$$\text{prob}(A \text{ e } B) = \text{prob}(A) \times \text{prob}(B|A) = \text{prob}(B) \times \text{prob}(A|B)$$

Disso podemos extrair a fórmula para o cálculo da probabilidade condicional:

$$\text{prob}(B|A) = \text{prob}(A \text{ e } B)/\text{prob}(A)$$

O teorema de Bayes, em sua versão mais simples, seria portanto (substituindo $\text{prob}(A \text{ e } B)$ por sua equivalente):

$$\text{prob}(B|A) = \text{prob}(B) \times \text{prob}(A|B)/\text{prob}(A)$$

O interesse desse teorema para uma explicação do papel da confirmação empírica se mostra quando queremos calcular a probabilidade condicional de uma hipótese (H) ser verdadeira dada uma evidência empírica (E). Neste caso, tomaríamos a fórmula $\text{prob}(H|E) = \text{prob}(H) \times \text{prob}(E|H)/\text{prob}(E)$. Ao interpretarmos isso, podemos perceber que $\text{prob}(H|E)$ é maior do que $\text{prob}(H)$, se $\text{prob}(E)$ for muito pequena, ou seja, se E for improvável, embora seja provável no caso de H ser verdadeira. Conseqüentemente, mesmo quando um cientista desconfia da improbabilidade de uma evidência, tendo ou não razões objetivas para isso, ele se veria compelido a alterar seu grau de crença, quando os indícios atuarem no sentido de tornar a probabilidade condicional da hipótese testada maior do que a probabilidade da mesma antes do teste. Isso vale matematicamente, portanto, quer para dados objetivos, quer para probabilidades subjetivas, isto é, quando um indivíduo atribui um valor à hipótese entre 0 e 1 arbitrariamente.

A $\text{prob}(E|H)$ é a medida da probabilidade de E ser verdadeira caso H seja verdadeira. Ela é mais alta, portanto, se as evidências obtidas corresponderem às previsões da hipótese e mais baixa, analogamente, se as

evidências não acompanharem as previsões. A $\text{prob}(H)$ é a chance de a hipótese ser verdadeira independentemente dos testes a serem realizados. Se a evidência (E) é um resultado surpreendente, a $\text{prob}(E)$, isto é, as chances de que E seja verdadeira independente da verdade da hipótese, será muito baixa.

Assim, suponhamos que um cientista tivesse pouca expectativa de que a hipótese da relatividade geral de Einstein (H) fosse verdadeira, considerando improvável a curvatura da luz quando próxima do sol. O cientista escolheria valores arbitrários para as probabilidades independentes de H e de E (sendo E a ocorrência da curvatura da luz na observação do eclipse total). Ele iniciaria supondo os seguintes valores arbitrários: $\text{prob}(H) = 0,1$ $\text{prob}(E) = 0,1$. A $\text{prob}(E|H)$ reflete o quanto a evidência correspondeu às previsões de H . Digamos, assim, que o fenômeno observado durante o eclipse aproximou-se em 80% dos números previstos por Einstein (o grau de curvatura, por exemplo), o que nos dá um valor 0,8. Aplicando esses valores à fórmula $\text{prob}(H|E) = \text{prob}(H) \times \text{prob}(E|H) / \text{prob}(E)$, temos: $(0,1 \times 0,8) / 0,1$. Isto é, $\text{prob}(H|E) = 0,8$. Agora o cientista pode mudar o seu grau de confiança na teoria da relatividade geral, uma vez que a confirmação detectada no eclipse observado elevou a probabilidade para 0,8. De acordo com os adeptos do bayesianismo, novos testes da teoria einsteiniana devem começar com a $\text{prob}(H) = 0,8$ e o valor obtido com nova evidência pode acrescentar ou diminuir a probabilidade condicional, de modo que a cada nova evidência o valor se ajuste e incorpore todos os testes já feitos. Digamos que seja possível testar outras previsões da teoria da relatividade geral e, neste caso, a evidência tenha se adequado apenas em 50% à previsão (isto é, $\text{prob}(E|H) = 0,5$). Tomando a evidência com $\text{prob}(E) = 0,7$ (a essa altura o cientista está apostando que a evidência tem mais chance de ocorrer do que de não ocorrer), temos agora que $\text{prob}(H|E) = 0,8 \times 0,5 / 0,7 = 0,57$. Tais valores significam que a ocorrência de uma evidência como a mencionada tornaria a teoria de Einstein um pouco menos provável do que antes do experimento. Nosso caso (semi-)hipotético mostra o suporte com que a teoria testada variou conforme o surgimento de mais dados empíricos, acompanhando sempre o que tais dados indicavam. Uma teoria pode ter uma probabilidade prévia 0,1 e depois ser favorecida pela evidência, mudando a probabilidade prévia para 0,8 em novos testes, caindo para probabilidade 0,5 e assim por diante.

A expectativa é que, com mais evidência disponível a probabilidade da

hipótese vá convergindo para um valor que variará muito pouco, refletindo igual convergência na opinião dos especialistas, para os quais os mesmos dados estão disponíveis, ainda que cada um parta de valores iniciais diferentes, refletindo seus preconceitos e diferentes expectativas. Daí que uma teoria altamente confirmada passe a gozar de ampla aceitação pela comunidade científica, de modo que, sem uma evidência muito determinante para derrubar tal confiança, sequer vale a pena questioná-la.

Há vantagens óbvias em assumir o bayesianismo. Uma delas é acomodar casos históricos de abandono de teorias em favor de descrições alternativas, já que anomalias proibidas por uma teoria seriam evidência em contrário a ela e, caso tais anomalias sejam previstas pela nova teoria, seriam evidência a favor desta última. No último caso, é de se esperar que a evidência já conhecida que confirmava uma teoria abandonada seja também evidência a favor da nova teoria, com a diferença que a probabilidade prévia da evidência antiga já seria alta (afinal já era bastante conhecida); são as novas evidências, antes desconhecidas, que aumentam a probabilidade condicional da nova teoria, já que elas ocorrem como a nova teoria prevê. Além do mais, o bayesianismo incorpora o aspecto crítico ressaltado pelo falsificacionismo, isto é, a característica da boa ciência em não se apegar dogmaticamente a uma teoria, mas prestar atenção aos indícios, ajustando o grau de crença a eles, sempre com abertura para mudança de opinião, com a ocorrência de novos dados. Mesmo quando falamos de probabilidade subjetiva (ou seja, arbitrária), o importante é que o cientista reconsidere seus graus de crença em relação ao cálculo bayesiano. Nestes casos, importa menos de quais os preconceitos o cientista parte, mas para onde a evidência o leva gradualmente.

Em queensem as vantagens do bayesianismo, alguns filósofos da ciência desconfiam da recomendação bayesiana de tomar o valor da $\text{prob}(H|E)$ como nova $\text{prob}(H)$ em testes futuros. O teorema em si não explica isso, que parece um passo adicional. O maior, problema, entretanto, é que, se aplicado ao problema da indução, a ideia de que haverá convergência entre diferentes agentes racionais só é verdadeira para alguns valores iniciais mais plausíveis. Pessoas seguindo as recomendações bayesianas, caso partam de valores mais extremos, não chegarão a valores finais convergentes. E não há explicação bayesiana para pessoas que acreditam, por exemplo, que a lei da queda livre de Galileu não está confirmada porque é possível que a natureza não seja uniforme.

Os bayesianos diriam que a pessoa é irracional, mas nesse caso não saberíamos porque ela é irracional (já que não há uma teoria bayesiana relacionando os valores iniciais subjetivos e a racionalidade). À dificuldade de justificar a uniformidade da natureza, proposta por Hume, soma-se a de explicar porque pessoas racionais deveriam esperar que a natureza seja uniforme.

2. Metafísica da ciência

As questões que compõem a metafísica da ciência são relativas ao modo como a realidade é ou não capturada pelas teorias científicas mais consensuais e ao modo como se compreendem conceitos que, embora não possam ser definidos cientificamente, são absolutamente importantes para compreender as afirmações científicas sobre o mundo natural. Uma parte das discussões metafísicas é sobre ontologia, isto é, sobre que tipos de seres existem. Na medida que há várias teorias científicas que lidam com partículas fundamentais e outras entidades que não podem ser verificadas por uma observação simples, a filosofia da ciência interessa-se em saber o status ontológico de tais entidades. Essa questão dá a tônica do debate em torno do realismo científico. Outra boa dose de questões metafísicas pode ser estabelecida pela pergunta: “o que é _____?”. No caso específico das questões derivadas da prática científica (teórica e experimental), recorre-se a noções tais como: leis, causas, modelos de explicação. O que são tais conceitos? Investigá-los é também tarefa da filosofia da ciência, já que saber sobre a existência de tais entidades usadas nas ciências é parte da metafísica da ciência.

2.1 Explicação científica, leis e causalção

Dentre os objetivos da ciência, um que parece especialmente relevante é o de fornecer explicações. Basta pensar nas perguntas que esperamos que os cientistas sejam capazes de responder. Boa parte de tais perguntas possuem a estrutura “por que _____?”. Por que a lua possui fases? Por que os dias são mais frios no inverno e mais quentes no verão? Por que o preço do gás de cozinha aumentou? Por que a superfície do lago está congelada? Por que aquele prédio desabou? Por que Fulano está com febre? Todas essas questões são típicas de um pedido por explicações. Espera-se que uma pessoa com

conhecimento de causa saiba justificar a ocorrência dos fenômenos que motivaram as perguntas iniciadas com “por que”, de modo que o ouvinte entenda como o fenômeno a ser explicado é uma ocorrência natural dos dados elencados. Por exemplo, para explicar o lago congelado, basta lembrar que a água se solidifica abaixo de 0°C e que a temperatura média da noite anterior foi de -14°C. Assim, era de se esperar que a superfície do lago congelasse. As temperaturas do inverno e do verão podem ser explicadas, por sua vez, pela inclinação do planeta que, ao orbitar em torno do Sol expõe um de seus hemisférios a uma distância menor daquele astro e o outro a uma distância maior, fazendo com que as estações sejam invertidas nos hemisférios diferentes e, assim, sugerindo que as mudanças de temperatura são aquilo que se espera em tais condições. As fases da Lua são explicadas pelas trajetórias já conhecidas do Sol, da Terra e da própria Lua; o preço do gás, pelas leis de mercado e situações particulares da economia atual; o desabamento do prédio, por leis da física em conjunto com a remoção de um pilar da construção por operários desavisados; a febre, pela ocorrência de uma infecção e pelo fato de o organismo liberar anticorpos para combater tal infecção, aumentando a temperatura corporal em 1 ou 2°C. Respostas a uma pergunta do tipo “por que _____?” assemelham-se à estrutura de um argumento dedutivo. O primeiro autor a ressaltar isso foi Carl Hempel, juntamente com Paul Oppenheim em 1948. Posteriormente, no livro *Aspects of scientific explanation* (1965) [Aspectos da explicação científica], Hempel desenvolveu seu modelo de explicação, o qual tornou-se, quer para complementação, quer para crítica, o ponto de partida básico para abordar a explicação científica. O modelo hempeliano nomeia o fato a ser explicado como *explanandum*. As leis científicas e as condições particulares que explicam o fato em questão compõem o *explanans*. Uma vez que a estrutura é de uma dedução, o que Hempel espera é que o *explanandum* seja uma consequência lógica do *explanans*. Uma das exigências de Hempel é que as condições e as leis sejam expressas em enunciados verdadeiros, de modo que garantam o *explanandum* assim como premissas verdadeiras garantem a conclusão de uma dedução válida. O modelo de explicação hempeliano é o seguinte:

C_1, C_2, \dots, C_k (Enunciados das condições antecedentes) L_1, L_2, \dots, L_r (Leis gerais)	<i>Explanans</i>
E (Descrição do fenômeno empírico a ser explicado)	<i>Explanandum</i>

Para tornar o esquema menos abstrato, tomemos um exemplo bastante simples de explicação científica onde 'P' é a pressão, 'V' o volume, 'T' a temperatura e 'k' uma constante. O fenômeno E a ser explicado é a diminuição do volume de um sistema fechado (por exemplo, diminuição do volume de um gás num frasco hermético). O experimento é conduzido de modo a manter a temperatura constante e aumentar a pressão. Tais são as condições iniciais. A lei da qual derivamos logicamente a explicação do fenômeno é a de Boyle, expressa na seguinte equação:

$$\text{Lei de Boyle: } P.V=k.T$$

Com tais dados, podemos explicar facilmente a diminuição do volume do gás no frasco: A lei de Boyle estabelece que o volume é diretamente proporcional à temperatura e inversamente proporcional à pressão. Dado que a temperatura permanece sempre constante e a pressão no frasco aumenta, resulta logicamente que o volume do gás diminui. Isto é, o fenômeno observado é coberto pela lei de Boyle, dado o aumento da pressão e da manutenção da temperatura.

Lei de Boyle ($P.V/T = k$) (Lei geral) Temperatura constante (condição específica) Aumento da pressão (condição específica)	<i>Explanans</i>
Diminuição do volume na mesma escala (fenômeno a ser explicado)	<i>Explanandum</i>

Tal modelo foi chamado, por razões óbvias, de modelo Nomológico-Dedutivo (D-N, de *deductive-nomological*) de explicação. Nomológico por partir de leis (*nomos* é lei, em grego) e dedutivo porque a conclusão é consequência lógica das leis e condições específicas. O modelo D-N também pode ser utilizado

para explicar a incorporação de uma lei menos geral num quadro teórico mais geral. Assim, a gravitação de Newton e a condição de que se esteja falando de um fenômeno que ocorre em uma altura muito pequena em relação à superfície da Terra explica porque a lei da queda livre de Galileu tem os resultados que tem. A relatividade geral explica os resultados de Newton, quando os fenômenos não possuem a velocidade próxima à da luz. Cada nova teoria abarca as anteriores, nos casos em que as teorias menos gerais acertam suas explicações. Em filosofia da ciência, a acomodação dos resultados de uma lei menos geral numa de maior generalidade é conhecida como redução. Muitos filósofos do positivismo lógico consideravam a redução de teorias uma das metas da ciência para um dia, talvez, possuímos uma ciência unificada.

O modelo D-N, no entanto, só se aplica a explicações determinísticas, isto é, àquelas em que se presume que as leis enunciadas no *explanans* sejam invariáveis e invioláveis, de modo que não haveria exceção às mesmas. E o fato é que há várias ciências nas quais este tipo de determinismo rigoroso não poderia ser estabelecido, ou não teríamos qualquer lei para enunciar. Consideremos casos da biologia ou da medicina e da farmacologia, nos quais há tanta complexidade de variantes, ou nos quais sabe-se de antemão que há apenas uma determinada correlação estatística muito forte. Hempel também sugeriu que seu modelo poderia ser estendido à explicação para fenômenos que só recebiam suporte estatístico do conhecimento científico disponível. Nestes casos, a regularidade estatística em conjunto com condições específicas tornaria provável o fenômeno a ser explicado. Como o *explanandum*, entretanto, não seria uma consequência lógica do *explanans*, mas apenas uma conclusão provável do mesmo, este tipo de explicação foi batizado de Estatístico-Indutivo (I-S, de *inductive-statistical*). Este modelo adaptado teria a seguinte formulação:

Leis probabilísticas	<i>Explanans</i>
Condições específicas	
Fenômenos a ser explicado	<i>Explanandum</i>

Assim, se queremos explicar o surgimento de um câncer no pulmão de um paciente, podemos recorrer às condições específicas de o mesmo paciente fumar e possuir antecedentes genéticos em conjunto com leis estatísticas de que o fumo aumenta a probabilidade de câncer no pulmão e de que condições

genéticas também aumentam tal probabilidade. Obviamente, tal explicação só mostra que o paciente tinha maior probabilidade de desenvolver o câncer que desenvolveu e não que o paciente necessariamente apresentaria a doença.

Muitos filósofos da ciência encontraram problemas nos modelos D-N e I-S de Hempel, preferindo tratar a ambos como modelos de explicação por leis de cobertura, já que, em síntese, o que a explicação hempeliana exige é que se encontrem leis capazes de cobrir os fenômenos a serem explicados ou as leis a serem reduzidas. Não está claro para os críticos, se o mundo apresenta a mesma simetria que o modelo de leis de cobertura sugere. O problema da simetria surge quando trocamos os papéis do *explanandum* com o de uma das condições específicas do *explanans*. Certamente a pressão pode ser explicada em termos de volume ou de temperatura como casos da Lei de Boyle. A lei dos gases perfeitos é adequada a tal simetria. Mas será que poderíamos explicar que o paciente seja um fumante devido ao fato de ter desenvolvido câncer? Um exemplo D-N famoso por explicitar que há algo de errado na simetria de tal modelo é o do mastro de um navio. É possível explicar o tamanho da sombra produzida pelo mastro, quando sabemos a altura do mastro e o grau de inclinação com que o raio do sol atinge o chão, acrescidos de leis geométricas. Intuitivamente, dizemos que a sombra é explicada pela altura do mastro e pela direção do sol. Ao meio-dia, a sombra será mínima e quanto mais o sol avança para se pôr, mais a sombra aumenta. Mas se perguntássemos o porquê de o mastro ter a altura que tem, seria legítimo dizer que é devido ao ângulo de incidência do sol e o tamanho da sombra, os quais acrescidos de leis da geometria têm como resultado a altura do mastro? Nossas observações comuns e uma intuição metafísica mais simples é a de que o mastro tem sua altura tal como é independente de haver sombra, ao passo que pensamos que a sombra do mastro só existe enquanto houver uma fonte de luz e o próprio mastro. Para evitar exemplos similares, talvez seja necessário lançar mão de uma ordem de relevância do *explanandum* para o *explanans*. Assim o mastro seria relevante para o tamanho da sombra, mas este não teria relevância para explicar o tamanho do mastro.

O modelo de cobertura por leis pode ser pensado com alguma afinidade com explicações causais, principalmente quando as leis científicas são entendidas de forma determinística, embora, como veremos mais à frente, também haja quem fale de causalção probabilística. Há toda uma tradição

metafísica de explicação por meio de causas que remonta a Aristóteles (século IV a.C.), filósofo que sugeriu que conhecer é conhecer as causas. No caso específico da filosofia aristotélica, supunha-se a existência de quatro tipos de causas (material, formal, eficiente e final). Hoje em dia, quando falamos de causas nosso pensamento se volta quase que exclusivamente à noção de causa eficiente, isto é, do conjunto de fatores que ‘produzem’ os efeitos a serem explicados. Causas finais seriam relativas aos objetivos, às finalidades ou às funcionalidades de um objeto sob escrutínio. Estas parecem muito adequadas para explicações sobre motivações de agentes racionais, já que uma pessoa pode agir para que um determinado estado de coisas desejado ocorra. Seria, assim, uma espécie de explicação da causa (eficiente) pelo efeito. No caso da biologia, muitas vezes recorre-se à função de um órgão ou o papel de determinada característica de uma espécie para explicar seu sucesso adaptativo. Explicações por causas finais também são chamadas de explicações teleológicas (do grego, *telos*=fim) e são bastante problemáticas no âmbito da filosofia da biologia ou da religião, devido ao seu uso para sustentar propostas criacionistas e afins (pode-se pensar que as funções ou finalidades presentes nos seres vivos sejam obra de um agente racional, no caso um criador divino). Uma resposta a tal investida da religião no terreno da explicação científica é a lembrança de que as características que favoreceram a adaptabilidade de uma espécie não surgiram para isso, mas simplesmente os indivíduos que a possuíam sobreviveram, ao passo que os indivíduos sem tal característica não. Colocar um elemento de inteligibilidade ou desígnio da natureza dos seres vivos seria uma adição desnecessária, portanto.

Quanto às causas eficientes, desde que Hume problematizou a indução, a vida dos filósofos da ciência que gostariam de manter a metafísica mais afastada de suas teorias ficou mais difícil. Hume, filósofo empirista, sugeriu que a noção de causa decorre da observação empírica de uma associação constante entre dois fenômenos consecutivos, de modo que a ocorrência do primeiro nos leve a esperar a ocorrência do segundo. De acordo com Hume, a fonte dessa crença é empírica e não racional. Não é racional porque nada do que sabemos racionalmente precisa ocorrer empiricamente para atestarmos sua veracidade, basta reconhecer que negar uma verdade racional leva a uma contradição, o que não ocorre quando pensamos que um efeito não se seguirá de uma causa. Por outro lado, só estamos em condição de alegar que A causa B depois de termos

visto B ocorrer na sequência de A por tantas vezes, que a mera presença de A nos leva a esperar B. Hume, entretanto, mostrou que a expectativa de que A cause B não pode ser, ela mesma, uma verdade empírica. Nela operam apenas o hábito derivado da conjunção constante e a crença de que isso continuará a ocorrer para sempre. Não podemos observar causas e, se formos empiristas como Hume, carregar a noção de causa com contornos de necessidade metafísica é suspeito. A noção de causalidade criticada por Hume era justamente a de que se A ocorresse, B necessariamente ocorreria. Verdades necessárias não seriam a meta da ciência empírica, que só lida com verdades contingentes. Verdades necessárias são as da matemática e da lógica, não da biologia, da química ou da física, e muito menos da economia ou das ciências sociais. Todas as leis da física poderiam ser pensadas como diferentes, sem que caíssemos em contradição. O exemplo que Hume dá é o de pensarmos que o sol pode não nascer amanhã, o que, por estranho que seja, não é contraditório, pois é até concebível sem violação de qualquer princípio do pensamento racional. Mas as verdades necessárias não podem ser negadas sem contradição: na geometria euclidiana um quadrado tem necessariamente 4 lados, e a soma dos seus ângulos internos é necessariamente 360° . Tente pensar num quadrado que não tenha 4 lados, ou cujos ângulos internos somem apenas 300° , ainda que consiga pensar tal figura, esta não será um quadrado. Um quadrado que não seja quadrado é contraditório. A proposta humeana para a causalidade pode ser deflacionada, metafisicamente falando, mas incorre em três problemas: o da distinção entre leis e regularidades acidentais, o da direção da causalidade e o da causalidade probabilística.

O primeiro problema é o de que nem toda regularidade é similar a uma lei, para fundamentar a relação causal. Há inúmeras regularidades deste tipo. É possível encontrar uma associação entre ser um livro sobre a minha mesa e ser meu livro, mas é apenas acidentalmente que um livro na minha mesa seja meu, ou seja, estar na minha mesa não causa a relação que este objeto tem comigo. Para este problema, uma tentativa de resposta é dada por F. R. Ramsey (1903-1930) e por David Lewis (1941-2001), assumindo que leis seriam generalizações verdadeiras que tomaríamos como pertinentes a um sistema ideal de conhecimento. A ebulição da água a 100°C decorre de leis sobre a fusão molecular, ao passo que estar na minha mesa e ser meu não faz parte de nenhum padrão de ciência aceitável. Outra linha de resposta é assumir que entre

a regularidade acidental e a relação causal há mesmo uma diferença metafísica, de modo que haveria uma necessidade natural da ocorrência do efeito caso ocorresse a causa, o que não é atendido pelos casos acidentais. O ônus dessa última resposta é retomar a ideia de necessidade, da qual Hume quis se afastar.

O segundo problema é similar ao do mastro no modelo de explicação por leis de cobertura. A mera conjunção constante não permite distinguir entre causa e efeito. Se A é regularmente acompanhado de B, como saber se não é B que causa A e não o oposto? Ainda que coloquemos o tempo como fator importante na distinção, teríamos o problema de confundir efeitos distintos de uma causa comum com uma relação na qual o primeiro efeito causa o segundo. O raio pode ser observado regularmente sendo seguido por um estrondo. Mas o primeiro não é causa do segundo. Antes, é a descarga elétrica que causa o clarão do raio e o estrondo do trovão. Uma explicação da direção da causalidade pela seta do tempo precisaria, para não ser circular, ser expressa em termos que não suponham, eles mesmos, uma seta do tempo. David Lewis também procurou resolver esse problema dizendo que a assimetria da causalidade é derivada da assimetria de “sobredeterminação”. É muito raro que eventos do presente sejam sobredeterminados por eventos do passado. Por exemplo, a morte de uma pessoa que levou um tiro e foi simultaneamente acertada por um raio. Já o presente ser sobredeterminado pelo futuro é algo bem mais normal, isto é, um evento presente deixa vários traços no futuro. Por exemplo, o disparo de uma arma faz com que haja impressões digitais, rastro de pólvora, o barulho, a perfuração do alvo, o suor no rosto do atirador etc. A não ocorrência do disparo sugere a não ocorrência de todos aqueles efeitos. Mas a ausência de um dos efeitos não implica a ausência da causa, desde que outros traços estejam ali. Hans Reichenbach (1891-1953) apresenta uma variante probabilística da postura de Lewis. Para o primeiro, as diferentes causas de um determinado efeito são independentes probabilisticamente falando. Já os diferentes efeitos de uma única causa possuem correlação estatística. Assim, tanto tomar café quanto fazer uso de certo tipo de antibiótico tornam provável o amarelamento dos dentes, mas isso não implica que pessoas que tomam café possuam mais chances de tomar aquele antibiótico. Por outro lado, como a falta de exercícios físicos pode levar à obesidade e a problemas respiratórios, é provável que pessoas obesas possuam problemas respiratórios. Os efeitos, para Reichenbach, são probabilisticamente dependentes; causas distintas, não.

O terceiro problema é o de saber se a causação precisa ser determinística (isto é, precisa que o efeito ocorra sempre que a causa ocorrer) ou se bastaria que a causa torne o efeito provável. Por muito tempo os filósofos tomaram a física como paradigma de ciência bem constituída. Deste modo, o determinismo presente nas teorias físicas era suposto na própria realidade e as áreas científicas que não fossem capazes de oferecer leis invioláveis pareceriam, de certo modo, imperfeitas ou imprecisas. Isso se modificou devido à teoria mais bem-sucedida da física, a mecânica quântica, que para ser desenvolvida exigiu o abandono do determinismo no nível microscópico. As equações da física quântica são tremendamente precisas, mas as leis que cobrem os fenômenos envolvendo partículas fundamentais são leis probabilísticas. Também podemos mencionar que as ciências da vida, com relevância sempre crescente, são dependentes de leis que descrevem tendências, disposições dos seres vivos, mas não uma necessidade imposta pelo livro da natureza. Isso para não falar nas ciências humanas, como a economia e toda a sua importância na condução dos governos etc. Construir um modelo de explicação probabilístico, entretanto, não é tarefa fácil. O modelo I-S de Hempel, por exemplo, só exige que as leis estatísticas impliquem maior probabilidade do *explanandum*. Mas quão alto deve ser o aumento da probabilidade do efeito para que o fator correlacionado conte como sua causa? Modelos mais recentes como o de Wesley Salmon (1925-2001), sugerem que basta que as causas aumentem a probabilidade de seus efeitos, não importando em que grau. Obviamente isso precisaria ser suplementado com ideias como as de Reichenbach, já mencionado, para não permitir que efeitos de causa comum sejam considerados como causa e efeito.

Mencionamos até aqui como algumas concepções de explicação científica sugerem que a natureza é regida por leis determinísticas em certos casos e probabilísticas em outros. Um modo comum de interpretar o modelo explicativo de cobertura por leis seria supor que as leis fundamentais da física e de outras ciências que possuem grande poder explanatório seriam uma expressão matemática da própria realidade. Essa interpretação pode seguir inferindo que, na medida em que os cientistas conseguem reduzir leis mais específicas a explicações mais e mais gerais, eles vão se aproximando de descobrir uma lei da natureza, que rege todo o cosmo, ou parte significativa de um conjunto de leis, que operam toda a realidade natural. As explicações baseadas no modelo de leis de cobertura dão a entender que um fenômeno

particular é causado por uma lei que rege casos similares, estando presentes certas condições. É assim, por exemplo, quando uma lei exprime uma força que condicione um corpo a se mover de um modo determinado. No caso da gravitação de Newton, a força (F) seria calculada pelo produto da massa de dois corpos (m_1 e m_2), inversamente proporcional ao quadrado da distância entre eles (r) e retificada por um valor constante (G):

$$\text{Lei da gravitação de Newton: } F=G.m_1.m_2/r^2$$

Com a lei da gravitação, é possível explicar e prever o posicionamento de um corpo celeste. Pode-se mesmo dizer que a força gravitacional produz o movimento do astro. Somos, portanto, levados a pensar que a Terra só se move em torno do Sol, porque uma força proporcional às massas dos dois corpos e inversamente proporcional ao raio da órbita terrestre em relação ao Sol atua para garantir o movimento do planeta. Mas há um ponto, ressaltado pela filósofa britânica Nancy Cartwright (1944 -), que não pode ser esquecido, ao interpretarmos tais leis fundamentais: todas as leis que sugerimos ser fundamentais só são verdadeiras se adicionarmos a condição de os demais fatores não cobertos pela lei em questão serem idênticos. O nome latino para isso é *ceteris paribus* e significa “todo o resto igual”. O texto de Cartwright, publicado em 1983 tem um sugestivo título: *How the laws of physics lie* [Como as leis da física mentem]. A ideia pode ser capturada com um caso mais simples: imaginemos a explicação para o palito de fósforo acender quando riscado. Nossa explicação dirá que a composição do fósforo é de material inflamável, que a cabeça do fósforo em atrito com a caixa produz uma faísca e que a faísca na presença do material inflamável torna o fogo uma ocorrência esperada. Quais condições além dessas precisam também ser observadas, para que o fósforo realmente se acenda? É preciso, por exemplo, que o palito esteja seco (o palito molhado não se acende) e que o fósforo não tenha sido riscado anteriormente (caso em que a cabeça deixa de ser inflamável). Numa explicação, quase sempre omitimos esses fatores também necessários para a ocorrência a ser explicada. Um outro caso similar é o da explicação de uma explosão. Embora, sem oxigênio, não seja possível a ocorrência de uma combustão, as explicações girarão em torno da presença de um combustível e de uma fonte de calor, já que o oxigênio é abundante no planeta. A lei da gravitação de Newton, para

Cartwright, também é verdadeira apenas se os objetos cujos movimentos são cobertos por tal lei não estiverem sujeitos a outros tipos de força e/ou fatores causais. Em outras palavras, trata-se de uma lei *ceteris paribus*. Assim como há uma lei para explicar a atração entre corpos com massa, há uma para explicar a repulsão/atração de cargas elétricas. O cálculo da força de atração (ou repulsão) entre cargas se dá pela lei de Coulomb, onde 'F' é a força de atração ou repulsão, 'q₁' e 'q₂' são as cargas que se interagem, 'r' a distância entre as cargas e 'k' uma constante determinada pelo meio:

$$\text{Lei de Coulomb: } F = k \cdot q_1 \cdot q_2 / r^2$$

De modo similar à lei da gravitação, a lei de Coulomb só seria verdadeira na desconsideração de outras forças que poderiam interferir. Na visão de Cartwright, a inclusão da cláusula *ceteris paribus* torna essas leis fundamentais menos interessantes, pois se aplicariam a eventos bastante simples ou idealizados e não aos casos reais, geralmente mais complexos. Para a autora, o mundo está cheio de fenômenos complexos, os quais tentamos reduzir e simplificar por meio de explicações. O que ocorre quando dois corpos possuem massa e carga? Nenhuma das duas leis sozinhas explica o fenômeno observado. Os físicos lançam mão de um artifício, que é fazer a soma dos vetores, descobrindo a resultante. Mas ao fazer isso eles estão violando a condição *ceteris paribus* daquelas leis. Estariam tomando, digamos, dois elétrons com a repulsão determinada pelo potencial de Coulomb e atração determinada pela gravitação universal como se elas atuassem em conjunto. O veredito de Cartwright é que apenas leis fenomenológicas, isto é, descritivas de fenômenos observados são realmente verdadeiras. Quanto mais essas leis descritivas são incorporadas em equações cada mais gerais, menos tais fórmulas serão verdadeiras. O preço a pagar por tamanha generalidade é justamente sacrificar a capacidade de descrever adequadamente a realidade (sobre a pertinência de considerar teorias verdadeiras ou falsas, ver o debate sobre o realismo científico na próxima seção). Para Cartwright, a explicação, tal como é praticada realmente por pesquisadores, envolve o caminho inverso do que parecem supor os defensores do modelo das leis de cobertura. Não é que as leis fenomenológicas sejam derivadas lógica e matematicamente das leis fundamentais "inscritas no livro da natureza", é o oposto que ocorre: a partir de leis fenomenológicas, são

construídos modelos idealizados e os mesmos permitem a elaboração de leis fundamentais bastante gerais e matematicamente eficientes, que não relatam exatamente o que ocorre no mundo, mas somente nos modelos a que se aplicam. Nesse sentido, tais leis não são entendidas como uma realidade por trás das aparências e dos fenômenos, mas apenas tentativas de sistematização e classificação de nossa diversidade de leis fenomenológicas. Os fenômenos muito particulares seriam representados como se fossem idênticos para permitir a construção matemática de um modelo, que nada mais é do que um simulacro para fins de previsões e generalizações.

2.2 Realismo e antirrealismo científico

Essa discussão sobre a capacidade de conhecimento da realidade além das aparências pela ciência é particularmente interessante para a questão do realismo científico e seus críticos. Em termos gerais, o realismo em filosofia remonta à tese metafísica de que o mundo exterior à mente é indiferente a (e independente de) nossas crenças, esquemas mentais, teorias e quaisquer conteúdos mentais que possuamos acerca dele. Essa posição contrastaria com o idealismo, segundo o qual o mundo é produto da mente, não existindo de modo independente. Dessa forma, o realista metafísico supõe que a realidade (em sua independência) é que faz com que crenças sejam verdadeiras ou falsas, na medida que elas exprimem tal realidade correta ou incorretamente. Dificilmente poderíamos falar de cientistas idealistas no sentido anterior. A bem da verdade, a questão do realismo científico não gira em torno da realidade observável. Uma grande parte dos filósofos da ciência admite que as teorias científicas nos dão uma boa ideia de como é o mundo, enquanto constituído de objetos e propriedades detectáveis aos nossos sentidos. Quando o assunto, entretanto, é saber se a ciência nos proporciona um vislumbre dos aspectos inobserváveis da realidade do mesmo modo como o faz com os fenômenos e entidades observáveis é que entra a polêmica sobre a qual trataremos agora.

Teorias científicas muitas vezes usam termos para designar partículas fundamentais, tais como prótons, fótons, elétrons, pósitrons, neutrinos. Também para designar seres de tamanho muito reduzido como bactérias e vírus, bem como organelas muito pequenas, como células, mitocôndrias e segmentos de DNA. Em alguns casos, falam de moléculas, que seriam compostas de átomos.

Todos estes nomes têm em comum o fato de que seus referentes, se existirem, têm dimensões tais que ultrapassam a capacidade humana de verificação sem aparatos auxiliando nossos sentidos. Em alguns casos, por exemplo, é necessário um microscópio. Os microscópios ópticos podem ajudar a detectar certas entidades biológicas por meio da luz (recorrendo a diferentes técnicas como fluorescência, radiação ultravioleta, polarização etc.), mas estes microscópios não poderiam “mostrar” partículas de luz, isto é, fótons. E microscópios eletrônicos, que produzem imagens por meio do bombardeio de elétrons, embora possam reproduzir a forma de objetos de dimensões aproximadas às de um átomo, não podem “mostrar” um elétron, nem objetos supostamente menores. Ademais, como já dissemos, cada um desses instrumentos possui uma complexidade que envolve, além de técnicas de mensuração e instrumentação, confiar nas teorias que fundamentam sua tecnologia. Um microscópio eletrônico, por exemplo, é construído de acordo com o que se espera ser o comportamento de feixes de elétrons descritos pela física quântica. Confiar em uma imagem produzida assim é, de certa forma, assumir que a física quântica nos dá uma descrição muito próxima do que seja a realidade física microscópica. E isso retoma o problema que gostaríamos de responder.

Os realistas científicos supõem, grosso modo, duas coisas (além, é claro, do realismo metafísico): em primeiro lugar, que as teorias científicas são verdadeiras ou falsas mesmo quando elas se referem a aspectos inobserváveis do mundo. Em segundo lugar, que temos conhecimento aproximado da realidade em todos os seus aspectos, observáveis ou não, quando tal realidade está descrita pelas teorias mais maduras e bem-sucedidas de que dispomos. Os antirrealistas científicos, em suas várias facetas, negam uma das teses realistas acima. Duas formas de antirrealismo que trataremos aqui seriam o instrumentalismo, que nega ambas as teses do realismo científico, e o empirismo construtivo, que nega apenas a segunda tese.

O instrumentalismo foi a postura adotada pelos positivistas lógicos. Para eles, os termos que designam entidades inobserváveis não passavam de uma construção lógica para abreviar uma série de procedimentos e medições verificáveis. As teorias nas quais tais termos aparecem não deveriam ser entendidas, portanto, como descrições de uma realidade inobservável independente. No máximo, tais teorias seriam instrumento útil para fazer

previsões e simplificar cálculos, o que explica o nome de tal postura. Na visão dos pioneiros do positivismo lógico, teorias que se referiam ao inobservável não deveriam ser interpretadas como verdadeiras ou falsas, elas sequer teriam sido concebidas para este fim. Há também os positivistas que, como Hempel, gostavam de ressaltar a necessidade de princípios de transposição para permitir falar de verificabilidade. Por meio dos princípios de transposição é possível reduzir termos teóricos a enunciados observacionais sem a carga metafísica que os primeiros possuem. Neste último caso falar que um elétron saltou de um nível orbital para outro é apenas um modo de dizer que houve uma liberação ou absorção de energia, sendo que um orbital é uma abreviação para uma quantidade fixa de energia suportada antes de outro “salto” (ou liberada por um salto). O instrumentalismo, estava demasiadamente comprometido com uma leitura empirista da ciência e, embora tenha importância histórica incontestável, não possui muitos seguidores atualmente.

Uma linha de defesa do realismo científico contra o instrumentalismo é dada pelo raciocínio segundo o qual não se comprometer com a tese de que a ciência nos dá descrições muito próximas da verdade sobre o mundo (observável ou não) é assumir que o sucesso da ciência é uma enorme coincidência. Em outras palavras, seria um milagre que as teorias mais maduras, confirmadas por um grande número de experimentos e muito consensuais fossem capazes de ser um prodígio explicativo e preditivo se não chegassem pelo menos perto de uma descrição correta da realidade. E uma vez que as chances são extremamente reduzidas para que o sucesso da ciência seja obra de um acaso de sorte, a melhor explicação para tal sucesso é a capacidade de elaborar teorias muito próximas de uma descrição verdadeira da realidade. Realistas científicos pensam que o sucesso que motiva essa forma de raciocinar não é exclusivo das partes teóricas referentes ao mundo observável, mas também daquelas partes referentes à realidade inobservável, uma vez que tais elementos seriam indissociáveis do restante da teoria. Por isso, ao falar de verdade (aproximada) de uma teoria bem-sucedida, está inclusa a ideia de que os termos que nomeiam coisas inobserváveis não são apenas um modo de falar. Vários desses termos para inobserváveis são peça fundamental nas teorias em que são supostos e, se elas são verdadeiras, as entidades inobserváveis devem possuir várias das características atribuídas a elas pelas teorias de sucesso das quais são parte. O argumento que acabamos de mencionar é chamado de

“argumento do sucesso”, “argumento do milagre” ou, mais apropriadamente, “argumento sem-milagre”. O argumento sem-milagre é uma forma de inferência à melhor explicação, pois assume que a melhor explicação para o sucesso científico é provavelmente a explicação correta. Neste caso, o realismo seria a explicação mais provável do que o milagre, não havendo outra opção entre as duas.

Uma resposta antirrealista ao argumento sem-milagre é a de que é possível enumerar uma série de teorias obsoletas que, no passado, exibiram o tipo de sucesso que o realista sugere só poder ser explicado pela verdade da teoria. Teorias como as do éter luminífero, do éter eletromagnético, do calórico e do flogisto, por exemplo, foram capazes de explicar uma gama bastante abrangente de fenômenos relativos ao comportamento da luz, do calor e dos processos de combustão e oxidação. Além disso, aquelas teorias já abandonadas também fizeram previsões surpreendentes, de modo que não se pode negar que tenham sido, a seu tempo, habilitadas para a crença realista. O fato histórico, porém, é que tais teorias vieram a ser substituídas e as entidades que postulavam passaram a ser consideradas ficções desnecessárias a partir das teorias que vieram depois. Se pudéssemos fazer uma indução a partir do fato de a história da ciência ser um cemitério de teorias e entidades, deveríamos concluir que provavelmente as teorias maduras de hoje sofrerão o mesmo fim das teorias maduras do passado. Não haveria ligação, portanto, entre sucesso, verdade e referência a não ser o desejo dos realistas de que estivéssemos convergindo para uma compreensão cada vez maior da realidade inobservável. O desejo, entretanto, não é evidência para o realismo. Este modo de criticar o realismo científico deveu-se sobretudo a um artigo de Larry Laudan intitulado *A confutation of convergent realism* (1981) [Uma refutação do realismo convergente]. Embora não tenha sido chamado assim por Laudan, esse uso da história da ciência contra o realismo científico ficou conhecido como “indução pessimista” ou, devido ao fato de ser uma indução sobre teorias, “metaindução pessimista”. Apesar do nome que já entrou para o vocabulário da filosofia da ciência, há outras formulações dedutivas do mesmo argumento seja como redução ao absurdo, seja como um *meta-modus tollens* pessimista. Neste último caso, o realismo teria como consequência uma série de teorias bem-sucedidas e supostamente verdadeiras que, entretanto, seriam abandonadas com o tempo. Há, porém, filósofos que rejeitam ambos os modos de formulação do argumento

pessimista. Segundo eles é questionável a viabilidade de se fazer uma indução sobre fatos históricos, já que agentes racionais, ao contrário de corpos e forças naturais, podem aprender com os erros e se comportar de modo diferente ao elaborar novas teorias ou aprimorar as já existentes. Igualmente, uma formulação dedutiva que quisesse derrubar o argumento sem-milagre com uma lista de teorias abandonadas pode não alcançar sua finalidade, devido ao fato de o argumento realista ser uma inferência à melhor explicação, isto é, um argumento não dedutivo. Alguns críticos da metaindução pessimista chegam a considerar que a contenção de Laudan é uma falácia da taxa de base. Esta se caracteriza pela omissão do universo circunscrito no argumento. Sem precisar quantas teorias totalizam esse universo, não saberíamos se os contraexemplos levantados são um número significativo ou se são meras exceções à regra.

Há também uma forma mais tenaz de antirrealismo, defendida pelo canadense de origem holandesa Bas van Fraassen, no livro *A imagem científica*, de 1980. Van Fraassen chamou sua proposta de empirismo construtivo. Em oposição ao antigo positivismo, o empirismo construtivo não pensa que enunciados contendo termos teóricos para as partes inobserváveis do mundo careçam de significado. Pelo contrário, seriam igualmente verdadeiros ou falsos. O problema seria nossa dificuldade fisiológica de verificá-los. A perspectiva de van Fraassen admite a existência de planetas e outros objetos macroscópicos que demandem instrumentos para serem vistos. É que se estivéssemos perto o suficiente, poderíamos atestar tais entidades sem ajuda de qualquer auxílio visual. Mas observar diretamente um elétron exigiria que fôssemos seres diferentes do que somos e, por isso, tal via de conhecimento está restrita para nós. Neste sentido, van Fraassen reconhece a possibilidade de que o mundo seja tal como propõem as teorias científicas, mas só o que podemos saber é como são os aspectos observáveis do mundo. Sobre os aspectos inobserváveis caberia assumir nossa limitação epistemológica, suspendendo o juízo acerca de como são as coisas neste nível. O objetivo da ciência não seria, portanto, oferecer descrições verdadeiras, mas apenas empiricamente adequadas. Para que uma teoria goze de aceitação entre cientistas, tudo o que é preciso é que ela “salve os fenômenos”. O empirismo de van Fraassen recorda o problema de Duhem-Quine (ver seção 1.2), mostrando que há muitas possíveis teorias alternativas que são capazes de acomodar os mesmos dados empíricos, diferindo-se apenas na quantidade e nos tipos de entidades inobserváveis

postuladas. Estes casos seriam exemplos de subdeterminação da teoria pelos dados. A preferência por um modelo ou outro de teorias alternativas igualmente consistente com os dados se daria por fatores pragmáticos e não estaríamos em condição de nos comprometer com a ontologia fundamental dessas teorias. De um modo muito incisivo, o empirismo construtivo contribuiu para minar a inferência à melhor explicação, isto é, a ideia de que a melhor explicação dentre as disponíveis para um conjunto de fenômenos é provavelmente explicação verdadeira. Van Fraassen propôs que não é possível garantir que a explicação verdadeira esteja entre as alternativas e, mesmo que isso fosse o caso, os realistas não conseguem estabelecer um nexo que ligue virtude explanatória à verdade. Pode-se esperar que uma teoria verdadeira explique com perfeição um conjunto de fenômenos, mas o inverso não se aplica. Uma teoria que explica com perfeição um conjunto de fenômenos pode não ser verdadeira, e não temos sequer critérios para decidir entre rivais subdeterminadas pelos dados. Se a teoria T_1 é idêntica a uma teoria T_2 , exceto por suas partes referentes ao mundo inobservável, seríamos capazes de determinar qual delas está mais perto da verdade? Se dissermos que é a teoria que postula menos entidades, temos que dar boas razões para pensar que o mundo tenha mesmo menos entidades, isto é, que a simplicidade seja uma marca das teorias verdadeiras. Mas não poderia ser o caso de que o mundo seja complexo e nossa preferência pela simplicidade seja de ordem estética, ou mesmo pragmática, ou ainda em razão de economia? Qualquer que seja o caso, caberia ao realista provar seu ponto, o que o empirista construtivo considera ainda não realizado.

Os realistas costumam responder às críticas acima, dizendo que, embora duas ou mais teorias possam ser empiricamente amparadas pelos mesmos dados, isso não significa que sejam igualmente bem-suportadas, havendo melhor ou pior suporte empírico, o que favoreceria uma escolha dentre outras possíveis e o aumento das chances de uma teoria ser mais próxima de uma descrição correta da realidade. Outra resposta, agora dirigida especificamente à crítica da inferência à melhor explicação é baseada no fato de os antirrealistas não negarem que tal processo inferencial leve à verdade no nível do observável. Então é como se o realista supusesse que os mesmos processos inferenciais sejam válidos para todos os domínios da realidade e o empirista construtivo pensasse que uma mesma forma de raciocínio levasse a garantias distintas, caso estejamos falando dos aspectos observáveis da

realidade ou dos inobserváveis. Neste caso, os realistas pensam que o empirista precisa dividir com eles o ônus da prova.

O empirismo construtivo de van Fraassen também tem uma explicação para o sucesso da ciência: à semelhança das espécies biológicas, teorias competem entre si, de modo que vão restando somente aquelas mais bem-adaptadas. Supor que a ciência é predominantemente uma prática bem-sucedida é colher um ou outro exemplo não representativo daquelas teorias sobreviventes, ocultando o fato de que a maior parte da atividade científica é de fracasso, tentativa e erro.

Em reação aos argumentos antirrealistas, uma estratégia por parte dos realistas foi procurar estabelecer graus de comprometimento diferentes a tipos diferentes de teorias e entidades. Isso significa que haveria também diferentes graus de realismo científico, de modo que certas ciências seriam mais sujeitas ao antirrealismo do que outras. Mesmo em relação a uma teoria específica, haveria elementos de maior ou menor susceptibilidade a dúvidas céticas. Consequentemente, haveria elementos inobserváveis de uma teoria sobre os quais se estabeleceria maior ou menor confiança de serem parte constituinte da realidade. Assim, as formas de realismo que emergiram em resposta aos críticos adotaram algum tipo de seletividade, oferecendo critérios para discriminar, numa teoria, os elementos sobre os que se deve ser realista e sobre os que se deve ser antirrealista.

Três grupos de realismo científico se destacam pela estratégia seletiva: o realismo explanacionista, o realismo de entidades e o realismo estrutural.

O primeiro, também conhecido como explanacionismo, enfatiza que o realismo deve ser direcionado apenas às teorias de sucesso surpreendente, isto é, aquelas que deram novas explicações para fenômenos conhecidos e fizeram previsões de fenômenos até então desconhecidos, mas que foram confirmados empiricamente. Esse tipo de realismo sofisticado ainda procurou apostar suas fichas na inferência à melhor explicação e, embora mais refinado, não afastou totalmente o problema dos novos sucessos do calórico, do éter, do flogisto e de dezenas mais de teorias que, embora capazes de cumprir seu requisito, foram trocadas por outras mais recentes.

O segundo grupo, do realismo de entidades, duvida que se possa estabelecer a verdade no nível do teórico, mas supõe que entidades passíveis de ser experimentalmente manipuladas, a ponto de produzir fenômenos

controlados, devam existir. O realismo de entidades evita insistir na inferência à melhor explicação (cujo foco é teórico) e prefere comprometer-se com as causas mais prováveis inferidas de nosso conhecimento experimental. As propriedades com as quais interagimos experimentalmente seriam "verdades caseiras" e entidades com tais propriedades são provavelmente reais, embora as partes mais teóricas de sua descrição estejam sujeitas a grande modificação e, até mesmo, substituição total da teoria. Isso ocorreu com os muitos modelos atômicos, modificados sem abalar a crença em átomos. O principal problema dessa abordagem é que, sem teorias, sequer haveria conhecimento das "verdades caseiras" e, nesse caso, o realismo de entidades não pode, de fato se diferenciar de um realismo comprometido com o teórico.

A última das famílias de realismo sofisticado é o realismo estrutural. Sua ênfase está nas relações matemático-formais das teorias científicas e não no aspecto descritivo das mesmas. Uma das variações do realismo estrutural sugere que podemos conhecer descritivamente apenas a realidade observável, ao passo que a única coisa que podemos conhecer da realidade inobservável é a estrutura formal. A outra variação estabelece que tudo que há (observável ou não) é estrutura. O que ambas têm em comum é que o inobservável só pode ser captado por suas relações estruturais. Dessa forma, apesar de uma teoria como a de Maxwell, que versava sobre o éter eletromagnético, ter sido substituída por uma mais adequada, deitando fora o conceito de éter como meio para propagação da luz, suas equações continuam servindo para calcular e prever fenômenos luminosos. As relações formais continuam existindo depois que a teoria é trocada. Os problemas com o realismo estrutural vão desde um comprometimento adicional com a existência de formas abstratas (um reino platônico onde relações existam independentes daquilo que relacionam, isto é, seus *relata*) quanto com o abandono do que parece ser uma das metas mais preciosas para um realista científico: a descrição de como o mundo é, e não apenas de como certas grandezas se relacionam.

Há ainda quem procurou dispensar o debate entre realismo e antirrealismo como improdutivo. Artur Fine, por exemplo, sugeriu a existência de um núcleo de crenças e recomendações compartilhadas entre realistas e antirrealistas científicos. Na visão de Fine, ambas as posições filosóficas aceitam os resultados da investigação científica, concordam que tais resultados são amparados por evidência empírica, que existam algumas entidades e que

algumas das proposições da ciência são verdadeiras. As diferenças, de acordo com Fine, estariam nos acréscimos metafísicos e epistêmicos que cada posição insiste em fazer àquele núcleo comum. Uma alternativa aos arroubos metafísicos e epistêmicos de antirrealistas e realistas seria adotar somente o núcleo duro admitido por ambas as posições. Essa via intermediária foi chamada por Fine de Atitude Ontológica Natural (cuja sigla, em inglês é NOA). Com o adjetivo “natural”, Fine sugere que não extrapolar o núcleo duro é o modo como os cientistas encaram a questão, já que estes não estão interessados seja nas concepções pragmáticas, convencionalistas ou instrumentalistas sobre a verdade (tal qual propõem os antirrealismos), seja numa versão mais robusta como a de verdade como correspondência com o mundo (proposta pelos realismos). A NOA de Fine está longe de ser a última palavra sobre o debate, tendo sido recebida com bastante crítica de cada lado. Talvez o maior problema da NOA seja caracterizar precisamente o que seja esse núcleo compartilhado sem assumir implicitamente algum dos compromissos do realismo científico ou de sua contraparte antirrealista. Além do mais, se Fine admite que a ciência oferece explicações e verdades, parece justo que filósofos queiram interpretar tais dados, conferindo-lhes algum significado, como sempre fizeram. Afinal, adendos metafísicos, epistêmicos e valorativos são o objeto mesmo da filosofia. Se não interessam a certos cientistas, é porque estes não têm condição de investigá-los cientificamente. Se, por outro lado, despertam interesse de uns e outros pesquisadores, é por estes possuírem veia filosófica. Quem poderia culpá-los?

Questões de revisão

1. Quais são os tipos de questões que ocupam os filósofos da ciência?
2. Explique, com suas palavras, o problema da demarcação e o que ele tem a ver com a epistemologia da ciência.
3. Qual é a postura do positivismo lógico quanto ao problema da demarcação?
4. Em que consiste o problema da indução e por que ele é importante para a filosofia da ciência?
5. Qual a resposta falsificacionista para o problema da indução?
6. Por que, para Popper, o critério de demarcação verificacionista falha?
7. Enumere e explique com suas palavras as críticas ao falsificacionismo.
8. Explique, com suas palavras, como o teorema de Bayes contribui para

- contornar o problema da indução.
9. O grau de crença numa hipótese aumenta com evidências antigas, para o bayesianismo? Justifique.
 10. O que é a distinção entre contexto da descoberta e contexto da justificação?
 11. Por que os filósofos historicistas recusam a distinção entre descoberta e justificação?
 12. Diferencie período de ciência normal do período de revoluções científicas.
 13. O que é incomensurabilidade? Por qual razão a crença na incomensurabilidade torna difícil falar de progresso científico?
 14. Que tipo de definição de ciência cabe numa filosofia historicista da ciência?
 15. O que é o modelo de explicação por leis de cobertura? Dê exemplos diferentes dos disponíveis no texto.
 16. Qual é o problema da assimetria da explicação e quais soluções podem ser dadas a este problema?
 17. Por que Hume recusa tomar a causalção como uma relação de necessidade?
 18. Quais seriam as possíveis soluções para o problema da seta do tempo?
 19. Como a probabilidade pode ajudar a distinguir causas e efeitos?
 20. Por que o modelo de leis de cobertura falha, para Cartwright?
 21. Realismo científico é o mesmo que realismo metafísico? Justifique.
 22. Explique as razões pelas quais, para o instrumentalismo, teorias sobre inobserváveis não são nem verdadeiras nem falsas.
 23. Qual a diferença entre instrumentalismo e empirismo construtivo?
 24. Exponha, com suas palavras, a explicação do realismo científico e do empirismo construtivo para o sucesso da ciência.
 25. Explique o problema de Duhem-Quine e o que ele significa para o debate do realismo científico.
 26. Por que a metaindução pessimista tem esse nome e qual o seu papel no debate do realismo científico?

Questões de discussão

1. Você concorda com a afirmação de que a filosofia da ciência não interessa a cientistas de verdade? Por quê?
2. É possível fornecer uma justificação não indutiva para a indução? Por quê? Caso negativo, seria necessário tomar a ciência indutiva como uma

- atividade irracional? Qual sua perspectiva sobre isso?
3. Você concorda que a ciência não passa de uma atividade dedutiva, de conjecturas e refutações?
 4. Mesmo que nenhuma hipótese fosse testada isoladamente, não é mais racional modificá-la, após uma refutação, do que modificar crenças de fundo e hipóteses auxiliares muito bem estabelecidas? Justifique sua resposta.
 5. A história da ciência falsifica tanto o verificacionismo quanto o falsificacionismo. Concorda? Por quê?
 6. Você concorda que a alquimia foi ciência em um certo período? Justifique.
 7. Qual sua opinião sobre a distinção entre contexto da descoberta e contexto da justificação? Trata-se de uma distinção artificial ou salutar? Justifique.
 8. Na sua opinião, o bayesianismo pode, de fato, auxiliar a escolher entre teorias rivais?
 9. É importante que a ciência chegue por redução a uma teoria mais geral ou esse é um ideal irrealizável? Justifique.
 10. Não existe problema da assimetria da explicação, a ciência não se interessa por causas, e por isso, a sombra explica a altura do mastro da mesma forma como a altura do mastro explica a sombra. Concorda? Por quê?
 11. Não existe causalidade probabilística. Concorda? Por quê?
 12. Você concorda com a ideia de que a regularidade é suficiente para estabelecer uma relação causal?
 13. Na sua opinião, qual a melhor forma de distinguir as causas de seus efeitos? Por quê?
 14. A ciência está longe de ser uma atividade de sucesso. Concorda? Por quê?
 15. Você considera o realismo científico mais plausível ou menos plausível do que o empirismo construtivo? Justifique.
 16. O argumento sem-milagre é uma inferência para a melhor explicação, isto é, um argumento não dedutivo; por não ser dedutivo, meros contraexemplos são insuficientes para derrubá-lo. Concorda? Justifique.
 17. Quando antirrealistas assumem que os mesmos processos inferenciais dão garantias diferentes para aspectos observáveis e inobserváveis da realidade, cabe a eles justificar seu posicionamento e não a realistas. Concorda? Por quê?

Sugestões de leitura

Há bons materiais de introdução geral à filosofia da ciência em português. **Sobre filosofia da ciência em geral**, uma ótima leitura inicial, completa e acessível é a *Introdução à filosofia da ciência*, de Alex Rosenberg. Um texto mais didático, que também pode servir para começar é *Ciência: conceitos-chave em filosofia*, de Stephen French. Uma última recomendação é o capítulo 'Filosofia da ciência' assinado por David Papineau, inserido no *Compêndio de Filosofia*, livro organizado por Bunin e Tsui-James. Esse compêndio é interessante por também conter também um capítulo sobre filosofia da biologia e outro sobre filosofia das ciências sociais.

Em relação aos temas de **epistemologia da ciência** vale a pena consultar: *Filosofia da ciência natural*, de Carl Hempel, *A lógica da pesquisa científica* e *Conjecturas e Refutações*, ambos de Karl Popper, bem como *A estrutura das revoluções científicas* de Thomas Kuhn e *Contra o método*, de Paul Feyerabend. Se houver interesse em conhecer alguns artigos que expõem a crítica de Feyerabend tanto à redução positivista quanto à noção de progresso científico de Popper, sugiro meus textos 'Feyerabend e a crítica às condições empiristas de redução interteórica' e 'Sobre o progresso científico: verossimilhança ou incomensurabilidade?'. Talvez o leitor também aprecie *O progresso e seus problemas*, de Larry Laudan, para uma alternativa ao falsificacionismo e o historicismo relativista. A perspectiva de Lakatos sobre o falsificacionismo pode ser lida no capítulo 'O Falseamento e a Metodologia dos Programas de Pesquisa Científica' e a compatibilidade de seu racionalismo com a história da ciência em *História da ciência e suas reconstruções racionais*. Uma leitura mais clássica é a das *Investigações sobre o entendimento humano*, de David Hume, que colocou muito bem o problema da indução e influenciou todo um debate sobre causalção. O capítulo 5 do livro *Philosophy of science*, de Martin Curd e J. Cover contém textos interessantes para o debate a favor e contra o bayesianismo (incluindo 'Why I'm not a bayesian' de Clark Glymour). Um destaque para o comentário dos próprios organizadores da obra, que constitui uma ótima introdução ao bayesianismo e seus problemas. Uma sugestão desses autores foi seguida por Alex Rosenberg na já mencionada *Introdução à filosofia da ciência*, cuja seção 5.4 procura explicar a definição de probabilidade condicional e as probabilidades prévias em termos de diagramas

de Venn.

O debate de **metafísica da ciência** que reage a Hume pode ser muito aprofundado com a leitura de *Causalidade e direção do tempo*, de Túlio Aguiar. Sobre explicação científica, caso o leitor tenha afinidade com a língua inglesa, vale a pena ler o artigo seminal de Hempel e Oppenheim, 'Studies in the logic of explanation' e o livro *Aspects of scientific explanation*, de Hempel. O livro *How the laws of physics lie*, de Nancy Cartwright apresenta uma crítica importante ao modelo de leis de cobertura e um diálogo interessante com o antirrealismo de van Fraassen. O debate entre antirrealismo e realismo científico pode muito bem ser aprofundado com os livros *A imagem científica*, de van Fraassen e *Representar e intervir*, de Ian Hacking. O primeiro expõe a base para o empirismo construtivo, a posição antirrealista mais desafiadora atualmente. O segundo apresenta os princípios de um realismo de entidades, uma tentativa de tornar viável o realismo científico diante das críticas de van Fraassen, fazendo dobradinha com o livro já citado de Cartwright. Em inglês, uma boa coletânea de textos sobre o debate realismo x antirrealismo é *Images of science*, organizado por Churchland e Hooker. Outro livro bastante completo sobre o tema é *Scientific realism*, de Stathis Psillos; nele o leitor encontrará uma proposta de realismo seletivo explanacionista (intitulada *divide et impera*). Publiquei também alguns textos mais recentes sobre o realismo científico que talvez valham alguma leitura, especialmente 'Algumas razões para levar a sério a metaíndução pessimista', com uma tentativa de resposta realista em 'Uma solução baseada no realismo experimental para dois argumentos pessimistas'. Meu artigo 'Notas sobre o problema do realismo científico' é também uma introdução ao tema e 'Uma proposta em dois passos para reabilitar o realismo experimental' é uma resposta às críticas comumente levantadas contra o realismo de entidades. Para uma defesa do realismo estrutural, ver o livro *The structure of the world*, de Stephen French e o artigo 'Structural Realism: the best of both worlds?', de John Worrall.

Há também fartos **recursos digitais** dentre os quais gostaria de destacar dois em português e um em língua inglesa. A primeira recomendação é *Revista Crítica na Rede* (seção de filosofia da ciência). Disponível em <<https://criticanarede.com/ciencia.html>>. O site, mantido pelo professor Desidério Murcho (UFOP) reúne uma série de traduções de artigos e de capítulos de textos valiosos, bem como resenhas e ensaios. Talvez o melhor conteúdo de introdução à filosofia em língua portuguesa disponível gratuitamente. A outra recomendação

é a revista *Principia: an international journal of epistemology*. Disponível em: <<https://periodicos.ufsc.br/index.php/principia>>. Trata-se de uma revista especializada em epistemologia ligada ao Núcleo de Epistemologia e Lógica da UFSC. Embora seja um periódico internacional, a revista publica também artigos em português e pode ser uma ótima fonte de aprofundamento nos temas da filosofia da ciência.

Para os que possuem bom nível de leitura em inglês, minha maior sugestão é a *Stanford Encyclopedia of Philosophy*. Disponível em <<https://plato.stanford.edu/>>. Organizada por Edward N. Zalta, é o projeto mais bem-sucedido de disponibilização de conteúdo filosófico gratuito e de qualidade em escala mundial. Difícilmente um assunto filosófico escapa da abrangência dessa enciclopédia, que é constantemente alimentada e seus verbetes periodicamente revisados pelos autores. Além disso, os textos não têm pretensão de serem concisos, ao contrário, procuram incorporar a máxima de que tudo o que é preciso para a compreensão de um verbete deve estar no próprio texto do verbete. Um número razoável de traduções de Stanford Encyclopedia of Philosophy para o português começou a ser disponibilizado, em formato de livros temáticos, através da Série Investigação Filosófica. Eles podem ser baixados gratuitamente em <<https://wp.ufpel.edu.br/nepfil/serie-investigacao-filosofica/>>. Trata-se de uma realização do Grupo de Pesquisa Investigação Filosófica e do Núcleo de Ensino e Pesquisa em Filosofia da Universidade Federal de Pelotas, bem como da editora da referida Universidade. Já há um livro publicado sobre lógica e outro sobre epistemologia e filosofia da ciência. Outro livro da mesma série dedicado à filosofia da ciência (ainda haverá mais!) foi organizado por mim e já está no prelo, podendo, em breve, ser consultado no sítio eletrônico já referido. Há, portanto, farto material de filosofia da ciência, sobre confirmação, explicação, causalção, método, valores na ciência, realismo científico entre outros.

Referências

- AGUIAR, T. (2008). *Causalidade e direção do tempo*: Hume e o debate contemporâneo. Belo Horizonte: Ed. UFMG.
- CARTWRIGHT, N. (1983.) *How the Laws of Physics Lie*. Oxford: Clarendon.
- CHURCHLAND, P. e HOOKER, C. (1985). *Images of science: essays on realism*

- and empiricism. Chicago: University of Chicago Press.
- CURD, M. e COVER, J. (1998). *Philosophy of science: the central issues*. New York: WW. Norton & Company.
- FEYERABEND, P. K. (2007). *Contra o método*. Trad. Cezar Augusto Mortari. São Paulo: Ed. UNESP.
- FRENCH, S. (2009) *Ciência: conceitos-chave em filosofia*. Trad. André Klaudat. Porto Alegre: Artmed.
- FRENCH, S. (2014) *The Structure of the World: Metaphysics and Representation*. Oxford University Press.
- HACKING, I. (2012). *Representar e intervir: tópicos introdutórios de filosofia da ciência natural*. Trad. Pedro Rocha de Oliveira. Rio de Janeiro: EdUERJ.
- HANSON, N. R. (1958). *Patterns of Discovery*. Cambridge University Press.
- HEMPEL, C. G. e OPPENHEIM, P. (1948) 'Studies in the logic of explanation'. *Philosophy of Science* 15 (2):135-175.
- HEMPEL, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press.
- HEMPEL, C. G. (1981) *Filosofia da ciência natural*. Trad. Plínio Sussekind da Rocha. Rio de Janeiro: Zahar.
- HUME, D. (2004). *Investigações sobre o entendimento humano e sobre os princípios da moral*. Trad. José Oscar de Almeida Marques. São Paulo: UNESP.
- KUHN, T. S. (2007). *A estrutura das revoluções científicas*. Trad. Beatriz Vianna Boeira e Nelson Boeira. 9.e.d. São Paulo: Perspectiva.
- LAKATOS, I. (1979). O Falseamento e a Metodologia dos Programas de Pesquisa Científica. In: LAKATOS, I. e MUSGRAVE, A. (Eds.) (1979). *A Crítica do Desenvolvimento do Conhecimento*. Trad. Otávio Mendes Cajado. Cultrix/EDUSP, São Paulo.
- LAKATOS, I. (1978). *História da ciência e suas reconstruções racionais*. Lisboa: Edições 70.
- LAUDAN, L. (2011). *O progresso e seus problemas*. Trad. Roberto Leal Ferreira. São Paulo: UNESP.
- OLIVEIRA, T. (2012). Feyerabend e a crítica às condições empiristas de redução interteórica. *Kínesis*, Vol. IV, nº 08, p. 123-134. Disponível em <<https://revistas.marilia.unesp.br/index.php/kinesis/article/view/4489/3301>> Acesso em 04 de julho de 2020.

- OLIVEIRA, T. (2014a). Sobre a questão do progresso científico: verossimilhança ou incomensurabilidade? *Síntese: revista de filosofia*. Belo Horizonte, v.41, n.129, pp.129-144. Doi: <https://doi.org/10.20911/21769389v41n129p129-144/2014>
- OLIVEIRA, T. (2014b). Algumas razões para levar a sério a metaindução pessimista. *Principia: an international journal of epistemology*, [S.l.], v. 18, n. 2, p. 269-29. Doi: <https://doi.org/10.5007/1808-1711.2014v18n2p269>.
- OLIVEIRA, T. (2017). Uma solução baseada no realismo experimental para dois argumentos pessimistas. *Veritas*. Porto Alegre, v. 62, n. 3, p. 595-623. Doi: <http://dx.doi.org/10.15448/1984-6746.2017.3.28684>
- OLIVEIRA, T. (2019). Uma proposta em dois passos para reabilitar o realismo experimental. *Kriterion*, Belo Horizonte, nº 144, p. 727-748. doi: <https://doi.org/10.1590/0100-512X2019n144121tlo>
- OLIVEIRA, T. (2020). Notas sobre o problema do realismo científico. *Investigação Filosófica*. Macapá, v. 10, n. 2, p. 47-59. Disponível em: <<http://periodicos.unifap.br/index.php/investigacaofilosofica/article/view/5494/pdf>> Acesso em 23 de junho de 2020.
- PAPINEAU, D. (2013). Filosofia da ciência. In: BUNNIN, N. e TSUI-JAMES, E. P. (orgs) (2013). *Compêndio de filosofia*. Trad. Luiz Paulo Rouanet. 4.ed. São Paulo: Ed. Loyola.
- POPPER, K. R. (2013). *A lógica da pesquisa científica*. Trad. Leônidas Hegenberg e Octavianny Silveira da Mota. 2.ed. São Paulo: Cultrix.
- POPPER, K. R. (2018). *Conjecturas e refutações*. Trad. Benedita Bettencourt. Lisboa: Edições 70.
- PSILLOS, S. (1999) *Scientific Realism: How Science Tracks Truth*. Routledge.
- ROSENBERG, A. (2013). *Introdução à filosofia da ciência*. Trad. Alessandra Siedschlag Fernandes e Rogério Bettoni. 2.ed. São Paulo: Ed. Loyola.
- RUSSELL, B. (2018). *Os problemas da filosofia*. Trad. Desidério Murcho. Lisboa: Edições 70.
- VAN FRAASSEN, B. C. (2007). *A imagem científica*. Trad. Luiz Henrique de Araújo Dutra. São Paulo: Editora UNESP: Discurso Editorial.
- WORRALL, J. (1989). 'Structural Realism: The Best of Both Worlds?' *Dialectica*, 43, pp. 99-124.

Filosofia da Física

Diana Taschetto
Thales Borrely

Naturalmente, de um texto introdutório sobre filosofia da física, como este que o leitor tem em mãos, espera-se, antes de tudo, uma resposta para a pergunta: “O que é filosofia da física?”. Talvez isso o decepcione, mas não temos uma resposta objetiva para ela. Mais do que isso, ninguém tem. Ao menos, ninguém tem uma que seja consensualmente aceita pelos acadêmicos da área. Como o leitor provavelmente sabe, a separação entre filosofia e ciências é algo relativamente novo na história da humanidade, ainda que um pouco dessa divisão já pudesse ser vislumbrado até mesmo na Grécia Antiga – nos textos gregos, questões relativas ao funcionamento da natureza eram frequentemente separadas de questões sobre tópicos mais amplos que hoje atribuímos à filosofia, por exemplo. A demarcação entre filosofia e ciência não é clara; muitas posições diferentes a respeito dela coexistem. Talvez a posição mais comum seja aquela que entende que a filosofia se limita a certos questionamentos totalmente alheios às ciências naturais (a questão sobre Deus provavelmente é o mais famoso deles). Outros defendem a antiga posição, que data ao menos da Grécia Antiga, de que é possível o filósofo utilizar somente a lógica e a racionalidade para acessar o mundo; os métodos das ciências seriam, assim, dispensáveis. Há ainda aqueles que entendem que o papel da filosofia é criticar a ciência a partir de observações externas, feitas com algum distanciamento. Nesta visão, o filósofo pode contribuir, organizando e esclarecendo os conceitos da ciência; ademais, ele pode racionalizar os métodos da ciência e apontar os melhores caminhos para o cientista.

Ao mergulhar nas mais fundamentais teorias da física, entretanto, o leitor se dará conta, esperamos, de que é muito difícil manter qualquer uma dessas posturas. Nesse contexto, a arbitrariedade na separação entre filosofia e

física salta aos olhos; e se percebe com facilidade que a fronteira entre as duas áreas é, no mínimo, muito opaca. Se o leitor puder tirar apenas uma lição deste capítulo, que seja essa. Esperamos que a breve apresentação que faremos a seguir sobre mecânica clássica, relatividade e mecânica quântica o auxilie nisso. Veremos nesses exemplos que as teorias da física são sempre fundamentadas sobre um conjunto de posturas filosóficas; as técnicas da ciência – a matemática, a estatística, a computação e a experimentação – não são suficientes para construir novas teorias. E essa via é de duas mãos. Os resultados da ciência podem moldar, delimitar e contribuir para o esclarecimento de vários problemas da filosofia. Ademais, as teorias e os resultados da física constituem um valioso conjunto de dados que pode munir os filósofos interessados em questões como “o que são teorias?”; “como teorias são construídas?”; “como os dados obtidos da experiência podem ser utilizados para inferências” etc.

Por fim, uma última ressalva antes de apresentarmos as questões da filosofia da física. A ausência de fronteira clara entre filosofia e física pode ser instigante uma vez que, ao se dar conta dela, o acadêmico passa a ter acesso a um maior arsenal para se digladiar contra os problemas que lhe interessam – por exemplo, as sofisticadas técnicas da ciência podem ajudar a detectar alguns dos dogmas escondidos em certas posturas filosóficas; ambas, filosofia e ciência podem se beneficiar em tentar compatibilizar seus resultados –, porém ela também traz alguns perigos. Em particular, tentativas de extrair interpretações e generalizações de cunho filosófico das teorias da física devem ser feitas com muito cuidado. Nunca se pode perder de vista que qualquer filosofia que saia de uma teoria física é resultado, antes de tudo, da filosofia utilizada para alimentar a teoria. Com essa ressalva feita, vamos à física.

1. Espaço, tempo, movimento

Questões sobre a natureza do espaço, do tempo e do movimento (cinemática) estão no coração da filosofia da física. Isso porque tais debates permeiam não só uma teoria física ou outra – como mecânica estatística, ou teoria quântica – mas toda a física: daí seu central papel na discussão filosófica/científica sobre a natureza da realidade. Note: não vemos, cheiramos, ouvimos, ou sentimos o espaço e/ou o tempo. E ainda assim, temos a impressão de que temos um acesso praticamente direto a eles: o espaço

parece vazio, uma espécie de “receptáculo”; o tempo, algo fluido, “que passa”. Porém, como veremos a seguir (embora de maneira muito resumida e simplificada), a física modifica – e drasticamente – as noções do senso comum sobre o que é o espaço e o que é o tempo. Daí o papel fundamental dos resultados científicos para as discussões filosóficas. A realidade não é o que parece.

1.1. Absolutos ou relativos? Leibniz vs Newton

O estudante iniciado na literatura filosófica sabe que as diferentes coisas existentes podem existir de diferentes maneiras. O amor, por exemplo, existe – mas existe de uma maneira diferente com que seu celular existe, ou com que as estrelas existem, ou com que a solução de determinada equação existe. De que maneira o tempo e o espaço existem?

É comum pensar no espaço como um receptáculo a ser preenchido (essa é a expressão usada por Platão no *Timeu*) ou como uma enorme sala vazia. Assim como o seu quarto, por exemplo, contém uma cama, um armário, e assim por diante, o espaço contém sua cama, seu quarto, sua casa, o planeta Terra, a galáxia, e todas as partículas que compõem a matéria do universo. A nossa intuição é a de que o espaço “contém” objetos, mas de um modo muito estranho: a espacialidade dos objetos parece *coincidir* com “partes” do próprio espaço (observe que essa noção de “conter” é diferente daquela com que uma caixa contém objetos; a analogia com o receptáculo, portanto, é limitada).

Algo similar parece ser verdade sobre o tempo. Temos a intuição de que tudo o que acontece, acontece no tempo; até nossos pensamentos – aos quais não atribuímos, percebe-se, espacialidade – devem acontecer em algum instante no tempo (sonhei X depois de adormecer em um instante de tempo t , e antes de acordar no instante $t + \Delta t$, onde Δt representa uma variação de tempo). De fato, temos a impressão de que todo e qualquer processo ocupa uma “porção” do tempo; o tempo, portanto – assim como o espaço – parece ser dotado de *extensão*. E sendo assim, como estamos acostumados com receptáculos vazios, é fácil conceber o espaço vazio (sem matéria), é fácil conceber o tempo vazio (sem processos)⁹⁶. Mas será isso verdade? Será que faz sentido falar em espaço

⁹⁶Para entender o que queremos dizer com “universo sem processos”, imagine que um

e tempo “vazios”? Se sim, por quê? Se não, qual a alternativa?

Este é o debate sobre o status ontológico do espaço e do tempo. Os aspectos centrais dessa discussão foram delineados no século XVII por Newton e por Leibniz; o tema é hoje discutido na literatura sob a égide “substancialismo” vs “relacionismo”. Podemos colocar o debate nos termos das seguintes perguntas: serão o tempo e o espaço reais (como celulares, quartos e estrelas)? Ou serão eles epifenômenos emergentes de propriedades de objetos?

Para Newton, o espaço e o tempo são tão reais quanto o livro que você está a ler agora. Vejamos por quê.

Na obra “Diálogos sobre os dois principais sistemas do mundo” (1632), Galileu Galilei propõe uma série de “experimentos” com o propósito de demonstrar o caráter inconclusivo dos argumentos apresentados pelos “filósofos aristotélicos” contra o movimento da Terra. No mais notório deles, Galilei chama a atenção para o seguinte fato: as conclusões às quais chega uma pessoa que está viajando, por exemplo, em um navio, *sobre os eventos que acontecem no navio*, não dependem do fato de o navio estar parado em relação à costa ou em movimento retilíneo uniforme. Um exemplo disso é o ponto de impacto de uma pedra que cai do topo do mastro do navio; ou, de maneira ainda mais pitoresca, a facilidade com que

as borboletas e as moscas continuarão seus voos indiferentemente para todas as partes, e nunca acontecerá que se concentrem na parte endereçada para a popa, como se estivessem cansadas de acompanhar o curso veloz do navio, do qual seriam separadas, por manterem-se no ar por longo tempo; e se, queimando alguma lágrima de incenso, produzísseis um pouco de fumaça, veríeis que ela se eleva para o alto e como uma pequena nuvem, aí se mantém, movendo-se indiferentemente não mais para esta que para

ser todo-poderoso, Deus, ou, se você preferir, o Monstro do Espaguete Voador, congele o movimento de todas as partículas do universo. Embora tudo no universo esteja, por conta da ação da deidade em questão, imóvel – inclusive relógios –, nossa intuição nos diz que o tempo “continuará passando”.

aquela parte. *E a razão de toda esta correspondência de efeitos é ser o movimento do navio comum a todas as coisas contidas nele [...]* nunca sentistes em vós a participação daquele movimento, a não ser quando a embarcação, batendo na areia ou esbarrando em algum obstáculo, parou, e que vós, juntamente com os outros passageiros, colhidos de surpresa, precipitaram-se com grande perigo (p. 267; 334, ênfase nosso).

Nesta passagem está delineado o chamado *princípio da relatividade de Galilei*, segundo o qual *as leis da física são as mesmas para todos os sistemas de referência que diferem um do outro por um movimento retilíneo uniforme*, ou por uma escolha de origem (x_0, t_0) no espaço $(x_0 \in \mathbb{R}^3)$ e/ou no tempo $(t_0 \in \mathbb{R})$.

Neste contexto, “sistema de referência”, ou “referencial”, tem o mesmo significado com o qual você provavelmente está familiarizado desde a escola, isto é, um sistema de coordenadas (ou de eixos) cartesianas com origem (ponto $(x, y, z, t) = (0, 0, 0, 0)$) arbitrariamente estabelecida; é o “laboratório” onde estudamos a realidade (com esta definição, o estudante é convidado a ler novamente o enunciado do princípio da relatividade acima). Chamamos “*referencial inercial*” o referencial que está ou parado, ou em movimento retilíneo uniforme, isto é, se movendo em linha reta com velocidade constante. Agora, observe o seguinte. Pelo princípio da relatividade, se você está em um referencial inercial, as leis da física não te dirão – você pode fazer o experimento que quiser – se você está parado ou em movimento (se isso não está claro para você, leia novamente a citação acima). Por exemplo, imagine que você está no navio de Galilei. Uma situação em que o navio se move em movimento retilíneo uniforme para a direita enquanto um amigo seu está parado na praia, *no que concerne às leis da física, é equivalente* à situação em que o navio está parado enquanto seu amigo e a praia viajam para a esquerda em movimento retilíneo uniforme. A simetria da situação não admite que falemos em movimento de maneira “absoluta”, por assim dizer: do navio, você dirá que está em movimento em relação a seu amigo; seu amigo, porém, pode dizer o mesmo sobre você. A física é a mesma!

A pergunta surge imediatamente: “movimento” é, então, algo que

depende do referencial (e portanto, do observador)? Newton responde: “não!”. Segundo Newton, o tipo de movimento que acima discutimos – o movimento *relativo* – não é verdadeiro. O movimento *verdadeiro*, diz Newton, é o *movimento absoluto*. Mas o que é isso? Ora, pois! Como tudo o que se move, se move em relação a alguma coisa, o movimento absoluto é definido em relação a *quê* o objeto se move, a saber, o *espaço absoluto*. Mas como sabemos que existe tal coisa, o “espaço absoluto”? Responde Newton: “através de experimentos!”.

Vejam. Newton observa que argumentos relacionais não dão conta de explicar o que chamamos, em física, de *forças inerciais*. Esse tipo de força envolve, por exemplo, *rotações* (não referenciais com orientações distintas; sistemas ativamente rotacionais). Imaginemos – o estudante agora é convidado a fazer o que chamamos em física de *experimento mental* – um universo onde há apenas um par de esferas idênticas paradas ligadas por uma corda; as esferas são impulsionadas de modo que o sistema esferas-corda inicia um movimento de rotação ao redor de seu centro geométrico. Não fosse pela corda, as esferas tenderiam a realizar um movimento linear e uniforme. A tensão (força centrípeta), contudo, faz com que as esferas girem ao redor do eixo vertical que passa pelo ponto O.

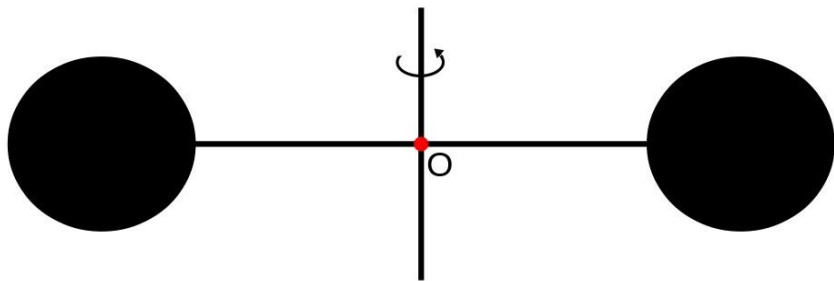


Figura 1. As esferas de Newton vistas de cima. A tensão na corda que as liga é evidência de uma rotação *absoluta* e de um referencial não-inercial.

A princípio, um observador sentado sobre uma das esferas, como descreve Galilei, pode achar que está parado. Contudo, o observador pode *medir* a tensão na corda; a tensão lhe permite perceber e quantificar a rotação; o movimento das esferas, portanto, *existe*. Não há como explicar a tensão na

corda apelando a argumentos relacionais; não há como dizer que o universo está girando e as esferas, paradas; se as cordas estivessem paradas, não haveria tensão. A única maneira de explicar a tensão é afirmando que as esferas estão girando em relação a um sistema de referência: o espaço absoluto.

No *scholium* do célebre *Principia Mathematica*, Newton propõe um experimento mental para demonstrar a existência do espaço absoluto que é, hoje, muito famoso: o chamado “experimento do balde”. O experimento tem algumas etapas: imagine, para começar, um balde de água suspenso por uma corda presa em algum ponto; imagine também, que o sistema corda/balde foi torcido muitas vezes, de tal modo que há energia potencial elástica o suficiente na corda para que o balde gire quando solto. O que você vê?

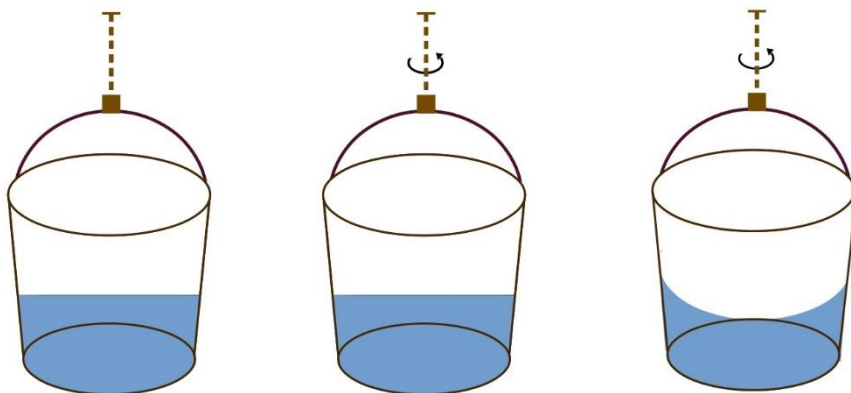


Figura 2. O balde de Newton. O movimento da água subindo pelas paredes do balde é evidência de uma rotação *absoluta* e de um referencial não-inercial.

Em um primeiro instante, a água está parada e sua superfície, plana; o balde também está parado. Em seguida, assim que o balde é solto e começa a girar, em um primeiro momento, a água no balde está parada; logo, porém, a água também começa a girar, e na medida em que gira, começa a subir pelas bordas do balde. Ora, pois, diz Newton! Isso é evidência de movimento absoluto! Pois, se o relacionista estivesse com a razão, a superfície da água teria a forma plana em ambos os casos, isto é, quando o sistema balde/água está parado e quando está girando, e seria plausível afirmar que a água e o balde estão em

repouso um em relação ao outro. Mas experimentos dizem diferente. Logo, o movimento do balde não é relativo, o movimento do balde é verdadeiro; e sendo que tudo o que se move, se move em relação a algo, o balde só pode estar se movendo em relação a uma coisa: o espaço absoluto.

Para mencionar agora um efeito inercial que não envolve rotações: imagine que você está sentado em um assento dentro de um avião prestes a decolar. Quando o avião começa a levantar voo, você sente uma força empurrando você para trás: esta força é um efeito inercial gerado por uma aceleração linear. Observe: você está *parado* em relação ao avião e, ainda assim, você sente uma força. Você, portanto, *sabe* que está em movimento.

Comparemos referenciais inerciais e não-inerciais: quando você estava no navio em movimento constante em linha reta, as situações “você se move em relação à praia” e “você está parado e a praia se move com velocidade constante em relação a você” eram equivalentes.

E no caso de referenciais não-inerciais? Observe: você, quando parado no avião, *sente uma força* quando o avião decola, uma pessoa que está na pista não sente essa força. A aceleração do avião quebra, portanto, a simetria entre os referenciais; não há como dizer que a situação em que o avião viaja para a direita e a pista está parada é equivalente ao avião estar parado e a pista viaja para a esquerda.

O estudante deve notar: *o que está em jogo aqui é a realidade do movimento*. É comum assumir que o movimento *existe*; afinal, assim nos parece. A realidade é mais complicada, porém – assim o estudante concluiu quando no navio de Galilei. Observe: Newton faz a distinção entre referenciais inerciais e não-inerciais – note que a distinção *não faz sentido do ponto de vista relacional* – porque a existência de *forças inerciais* leva Newton a concluir que *há referenciais não-inerciais*; e como explicar referenciais não-inerciais, *se não conseguimos explicar, antes, referenciais inerciais*? A *distinção* entre tipos de referenciais *deve existir* porque aprendemos, *com experimentos*, que em alguns referenciais não podemos determinar se estamos ou não em movimento, mas em outros, sim. *E a realidade do movimento, que provamos existir a partir de efeitos observados de forças inerciais, implica na realidade de um lugar onde as coisas se movem, implica na realidade dos instantes/intervalos de tempo em que o movimento ocorre*.

O espaço como entidade independente, portanto, existe. O tempo,

também.

Refleta sobre como as noções filosóficas (noções *substantivistas*) de espaço e de tempo de Newton condicionam o entendimento da mecânica clássica sobre o conceito de referencial.

Passemos às objeções. O mais proeminente crítico da teoria de Newton foi o filósofo alemão G. W. Leibniz. Para ele, as noções de espaço e tempo absolutos eram – como o estudante verificará em seguida – monstruosidades metafísicas. Para que possamos entender seus motivos, revisemos rapidamente os princípios filosóficos que servem de eixo ao seu pensamento, a saber, o Princípio da Identidade dos Indiscerníveis (PII) e o Princípio da Razão Suficiente (PRS). O PII diz o seguinte:

Se x e y compartilham todas as suas propriedades (que representaremos aqui por F), então x e y são idênticos (em símbolos, $\forall F |Fx = Fy| \rightarrow x = y$). Ou, reciprocamente, não há distinção sem diferenças.

Isso pode parecer simples, mas há várias sutilezas por trás do uso do PII. Por exemplo: que propriedades contam como F ? Propriedades qualitativas apenas? Coordenadas no espaço e no tempo? Ademais, o que conta como x e y ? Objetos? Possibilidades? Mundos? Agora, enunciemos o PRS:

O PRS diz que para todas as coisas que são o caso, deve existir um motivo, ou explicação, para o caso ser este, e não outro.

Por exemplo: se você consegue ver sua imagem quando diante de um espelho, mas não quando diante de uma parede de concreto, há um motivo para que seja assim; se você, estando no nível do mar, aquece água para preparar café, e a água vira gás a 100 C°, mas não a 80 C°, isso acontece por alguma razão, não por um capricho da Natureza. O estudante deve notar que, para Leibniz, o PRS é uma lei metafísica, isto é, aplica-se a todos os aspectos da realidade. Portanto, se todo objeto M existente no universo está, como afirma Newton, localizado, a cada instante t do tempo (absoluto), em determinado ponto

x do espaço (absoluto), então, deve haver um *motivo* para que, no tempo t , M esteja em x e não em x' ; ou ainda, em x no instante t , e não no instante t' . Em termos simples, a existência do espaço (tempo) absoluto implica na realidade da posição espacial (temporal) dos objetos contidos no espaço (tempo).

Porém, não há razão suficiente para a localização de objetos no espaço (tempo).

Logo, diz Leibniz, não existe algo como “localização dos objetos no espaço (tempo)”.

Logo, diz Leibniz, não existe algo como “espaço (tempo) absoluto”.

Analisemos a mesma questão à luz de PII. Lembre-se do princípio da relatividade: as leis da física são as mesmas em todos os referenciais inerciais. Lembre-se do que isso significa: não há como distinguir experimentalmente qual das situações – referencial parado ou em movimento constante em linha reta – é a “verdadeira”: se você está em um ponto no espaço-tempo (x, y, z, t) e seu amigo em $(x + a, y, z, t + b)$, $a \in \mathbb{R}^3$, $b \in \mathbb{R}$, as leis da física se aplicam igualmente (dizemos que as soluções das equações da teoria são ligadas por simetrias; neste caso, as chamadas “transformações de Galileu”, que envolvem translações, rotações, e combinações disso).

Recordamos o estudante: é um fato básico em física ser possível descrever qualquer evento idealmente pontual e de duração instantânea por uma coleção de quatro números que especificam sua posição espacial e seu instante no tempo, medidos em algum referencial. Sob a hipótese de que a mecânica newtoniana – e o espaço e o tempo absolutos por implicação – é verdadeira, considere as coordenadas (x, y, z, t) que localizam a ponta de seu nariz no instante em que você lê estas linhas (este será o evento que estudaremos). Observe: existem *infinitas possibilidades* – diferindo apenas por translações, rotações, ou combinações disso – de coordenadas espaçotemporais para a ponta de seu nariz. Do ponto de vista metafísico, cada uma dessas coordenadas é inteiramente diferente das outras. Porém, os diferentes pontos do espaço absoluto (identificado matematicamente com \mathbb{R}^3) e os diferentes instantes do tempo absoluto (identificado matematicamente com \mathbb{R}) não podem ser distinguidos empiricamente se toda a matéria no universo – ou ainda, apenas a

régua e o relógio utilizados para determinar as coordenadas espaçotemporais da ponta de seu nariz – forem trasladadas ou rodadas (e/ou o relógio adiantado ou atrasado) igualmente no espaço e no tempo!

Reiteramos: as possíveis localizações da ponta de seu nariz (ou de qualquer outro objeto), por conta da simetria euclidiana de \mathbb{R} e \mathbb{R}^3 são *fisicamente indistinguíveis*. Sob o ponto de vista de Newton, o mundo em que a ponta do seu polegar está em x no instante t é diferente do mundo em que a ponta do seu polegar está em $x' = x + 5$ no instante $t' = t - 2$, mas como leis da mecânica são invariantes por transformações de Galileu, essas diferenças não se deixam mostrar na experiência sensível. Lembre-se do enunciado do PII: não há distinção sem diferenças. Não nos surpreende, portanto, que Leibniz tenha criticado a mecânica clássica tão duramente. Observe que essa teoria – em total contradição com PII e PSR – gera um número infinito de mundos possíveis fisicamente indistinguíveis! Qual desses mundos seria “o verdadeiro”?

Recapitemos. As simetrias do espaço e do tempo absolutos levam a estruturas empiricamente inacessíveis; essas estruturas são incompatíveis com PII e PRS. O estudante deve agora se perguntar: se por hipótese, o espaço e o tempo não são absolutos, qual seria, então, a alternativa?

Vejam. Considere dois eventos pontuais quaisquer x e y de duração instantânea. É fácil ver que há uma relação temporal entre x e y : x acontece antes de, simultaneamente a, ou após, y . Em termos mais gerais, falemos de *intervalos*: x está separado de y no tempo por algum intervalo Δt que pode ser nulo, positivo ou negativo. A ideia de Leibniz sobre o tempo é essa: tempo se resume à *coleção de relações temporais entre eventos*. Observe: na ausência de eventos, ausentes suas relações, e por implicação, *ausente o tempo*.

Analogamente para o espaço. Os objetos estão a certas distâncias uns dos outros; isso define uma relação entre eles. A coleção de todas as relações espaciais entre os objetos existentes a cada instante no universo define o espaço. Portanto, ausentes os objetos, ausentes as relações espaciais entre eles, *ausente o espaço*.

O estudante é agora advertido: não obstante nossa intuição seja hoje, moldada à *la Newton* (pois afinal, aprendemos física clássica na escola), à época de Newton, *a intuição era outra, a saber, a relacional*, advogada por Leibniz. Isso significa dizer: se hoje a visão relacional parece estranha ao estudante, no século XVII, a visão de Newton soava herética aos ouvidos de seus

contemporâneos. Ressaltamos: não obstante as inteligentes objeções ao espaço e ao tempo absolutos apresentadas por Leibniz contra Newton, Leibniz defendida o *status quo*. A filosofia relacional do espaço e do tempo não é de autoria de Leibniz. Seu autor é Aristóteles.

Vale a pena falar um pouco mais a respeito. Para Aristóteles, tempo e espaço eram “qualidades secundárias” – não entidades – emergentes das relações entre os objetos. “Movimento” não era algo a ser descrito matematicamente (a matematização do tempo, espaço e movimento começou com os trabalhos de Kepler, Tycho Brahe e Galilei; atingiu seu ápice com a geometria analítica de Descartes e a física de Newton). O estudante deve observar o contraste filosófico que há aí: as categorias aristotélicas de “potência” e “atualidade” descreviam o “porquê” do movimento; “a pedra cai porque no chão é seu ‘lugar natural’; logo, a pedra tem ‘potência’ para cair; ‘movimento’ é a *passagem* da ‘potência’ para a ‘atualidade’”. Essa descrição qualitativa do movimento foi, com Kepler, Galilei, Descartes, e Newton, substituída por outra, *quantitativa*: queremos saber *como* os objetos se movem. Pense um pouco: você consegue fazer sentido da situação física indicada na figura 3 – você provavelmente já desenhou várias vezes algo semelhante quando na escola – em termos relacionais? Provavelmente, não. Sempre que desenha um esquema como esse, você imagina a posição e o movimento do corpo em relação aos eixos x e y (ou seja, em relação ao espaço absoluto). Repare, portanto, que até mesmo a física que você aprendeu na escola e provavelmente aplicou no vestibular pressupõe uma metafísica.

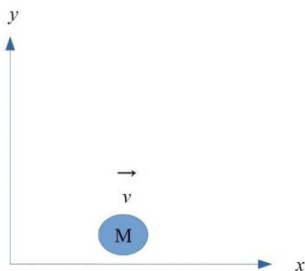


Figura 3. Representação no plano cartesiano de um corpo de massa M que se move com velocidade constante v na direção x .

Talvez o estudante já tenha ouvido falar da famosa frase de Newton “*hypothesis non fingo*” (eu não faço hipóteses). Para que não haja mal-entendidos, é importante enfatizar que, para Newton, *o postulado do espaço e tempo absolutos não eram hipóteses metafísicas* (embora fossem). Newton estava convencido de que *experimentos* (como o do balde) são *prova* da existência do movimento absoluto; a existência do espaço e do tempo absolutos seguem por implicação. De fato, o relacionista não consegue dar conta, com seus argumentos, de explicar o efeito de forças inerciais em corpos. Mas o fato de movimentos não-acelerados serem empiricamente inconsequentes incomodava os cientistas; e tendo a autoridade de Newton substituído a de Aristóteles, havia, de fato, a esperança, até o início do século XX, de eliminar esse pequeno problema por meio de experimentos óticos. O famoso experimento de Michelson Morley, do qual você talvez já tenha ouvido falar, tinha justamente esse objetivo: determinar o estado de repouso em relação ao espaço absoluto estudando o comportamento da luz. O resultado *negativo* desse experimento – que os cientistas, descrentes, repetiram várias vezes – corrobora as ideias de um jovem físico alemão que, em 1905, provou, com argumentos muito simples, que o princípio de relatividade de Galileu, aplicado às leis do eletromagnetismo, falseiam a metafísica newtoniana. Esse físico se chamava Albert Einstein.

Atualmente, à fábrica da realidade chamamos “espaço-tempo”. O estudante não deve interpretar o comentário como dizendo que a visão adequada é a relacional, de Leibniz e Aristóteles, e não a substantivista, de Newton. A questão é mais complicada. A realidade é muito mais plástica do que suspeitávamos. Veja só que bacana.

1.2. De espaço-e-tempo para espaço-tempo

A base física da relatividade especial, proposta na introdução do artigo “Zur Elektrodynamik bewegter Körper” (1905b) de Albert Einstein, é a seguinte:

Em todos os sistemas de coordenadas em que são válidas as equações da mecânica, também são igualmente válidas leis ópticas e eletrodinâmicas da mesma forma [...]. Vamos erguer à categoria de postulado essa nossa

suposição (a cujo conteúdo chamaremos daqui em diante “princípio da relatividade”); e além disso, vamos introduzir o postulado – só aparentemente incompatível com o primeiro – de que a luz, no espaço vazio, se propaga sempre com uma velocidade determinada, independente do estado de movimento da fonte luminosa. Esses dois postulados são suficientes para chegar a uma eletrodinâmica de corpos em movimento, simples e livre de contradições, baseada na teoria de Maxwell para corpos em repouso. (p. 891-892)

É isso que precisamos entender.

Maxwell provou ser consequência das equações da eletrodinâmica a velocidade da luz em um determinado referencial inercial ser c (em SI, 299 792 458 m/s). O estudante é agora recordado: as transformações de Galilei relacionam as coordenadas de um evento pontual de duração instantânea em um referencial em repouso $S(x, y, z, t)$ com as coordenadas (x', y', z', t') de um referencial S' que se move em velocidade constante v em linha reta na direção x relativamente a S (com suas posições coincidindo em $t = t' = 0$) da seguinte maneira:

$$x' = x - vt$$

$$y' = y$$

$$z' = z$$

$$t' = t.$$

Observe que isso implica na velocidade da luz em S' ser $c' = (c - v)$. Logo, $c' \neq c$, em contradição com o princípio da relatividade. Portanto, por *modus tollens*, uma das seguintes premissas estão erradas (ou ambas): (a) a velocidade da luz é constante em todos os referenciais inerciais; (b) as transformações de Galileu relacionam adequadamente as coordenadas espaço-temporais de referenciais inerciais S e S' . (a) é um resultado teórico bem corroborado experimentalmente; logo, devemos abandonar a premissa (b). De

fato, como veremos na sequência, a Natureza obedece não às transformações de Galileu, mas as chamadas “transformações de Lorentz”.

Iniciamos a discussão sobre as “transformações de Lorentz” definindo um conceito chamado “intervalo”. Vejamos. Sejam $X = (t_0, x_1, x_2, x_3)$ e $Y = (t_1, y_1, y_2, y_3)$ coordenadas de dois eventos quaisquer no espaço-tempo (identificados matematicamente com \mathbb{R}^4) localizados no *mesmo* referencial inercial. Definimos o *intervalo* entre X e Y como sendo a quantidade

$$I(X, Y) := I(X - Y) := c^2(t_0 - t_1)^2 - (x_1 - y_1)^2 - (x_2 - y_2)^2 - (x_3 - y_3)^2. \quad (1)$$

Logo ficará claro o porquê desta definição. Não perca de vista que ela se aplica a quaisquer dois eventos em um dado referencial.

Utilizemos agora a definição de velocidade (a velocidade escalar comum, aquela que você aprendeu na escola) para um feixe luminoso, isto é,

$$\text{velocidade da luz} = \frac{\text{distância percorrida pela luz}}{\text{intervalo de tempo}}. \quad (2)$$

Apliquemos (2), utilizando a mesma notação de eventos X e Y, isto é, imaginemos que o evento X é a partida da luz – em outras palavras, a luz é emitida no instante t_0 no ponto do espaço (x_1, x_2, x_3) – e o evento Y é a chegada da luz – em outras palavras, o feixe atinge a coordenada espacial (y_1, y_2, y_3) no instante t_1 . Lembrando que, dados dois pontos no espaço, a distância entre eles é a raiz da soma dos quadrados das diferenças entre cada uma das coordenadas que determinam os pontos. Para eliminar a raiz na fórmula da distância, apliquemos (2) tomando o quadrado dos dois lados:

$$c^2 = \frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}{(t_0 - t_1)^2}. \quad (3)$$

Passando o denominador para o lado esquerdo da equação, temos que o quadrado da distância percorrida pela luz é igual a $c^2(t_0 - t_1)^2$. Agora, podemos entender a definição (1): ela é a subtração entre o quadrado da distância percorrida pela luz entre os instantes t_1 e t_0 – nos quais ocorrem os eventos X e Y – e o quadrado da distância entre os pontos do espaço (x_1, x_2, x_3) e (y_1, y_2, y_3) – onde ocorrem os eventos X e Y. Em outras palavras, a noção de

intervalo serve para comparar a distância espacial entre dois eventos com a distância máxima que a luz (ou qualquer outra coisa) poderia percorrer entre os instantes de tempo que separam os eventos. Com isso é possível estudar se os eventos podem estar causalmente conectados.

Antes de debatermos causalidade, uma observação: pelos postulados de Einstein, em um outro referencial inercial S' , as coordenadas da partida e da chegada da luz na equação (3) seriam outras; a velocidade da luz, contudo, seria a mesma. Escrevendo as coordenadas do evento X e Y no referencial S' genericamente como $X' = (t'_0, x'_1, x'_2, x'_3)$ e $Y' = (t'_1, y'_1, y'_2, y'_3)$, temos:

$$c^2 = \frac{(x'_1 - y'_1)^2 + (x'_2 - y'_2)^2 + (x'_3 - y'_3)^2}{(t'_0 - t'_1)^2}. \quad (4)$$

Note que, tanto em (3) como em (4), se passarmos o denominador e, em seguida, os demais termos do lado direito para o lado esquerdo da equação, resgataremos a definição de intervalo de modo que $I(X, Y) = I(X', Y') = 0$. Aqui percebemos outra relevância do conceito de intervalo: *ele é invariante, i. e., é o mesmo em quaisquer referenciais inerciais*. Vamos entender agora o que significa um intervalo ser ou não ser nulo a partir de um sistema de classificação.

Classificação de intervalos. Em um referencial, o intervalo entre dois eventos diferentes X e Y são ditos ser

1. Do tipo *luz* de $I(X, Y) = 0$;
2. Do tipo *tempo* se $I(X, Y) > 0$;
3. Do tipo *espaço* se $I(X, Y) < 0$.

Portanto,

- A) se $X = (t_0, x_1, x_2, x_3)$ e $Y = (t_1, y_1, y_2, y_3)$ são do tipo luz,

$$c^2 = \frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}{(t_0 - t_1)^2}.$$

Neste caso, pode existir um sinal luminoso que se propaga com

velocidade c e que conecta X e Y.

B) Se $X = (t_0, x_1, x_2, x_3)$ e $Y = (t_1, y_1, y_2, y_3)$ são do tipo tempo,

$$c^2 < \frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}{(t_0 - t_1)^2}.$$

Neste caso, pode existir um sinal que se propaga com velocidade menor que c e que conecta X e Y.

C) Se $X = (t_0, x_1, x_2, x_3)$ e $Y = (t_1, y_1, y_2, y_3)$ são do tipo espaço,

$$c^2 > \frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}{(t_0 - t_1)^2}.$$

Neste caso, só poderia existir um sinal que conecta X e Y se este se propagasse com velocidade maior que c , o que é proibido pelo postulado de Einstein.

Chamamos a atenção novamente para o fato de que é muito bem corroborado experimentalmente que a velocidade da luz é a velocidade limite da natureza: jamais foi observado um sinal de qualquer espécie se movendo com velocidade maior que a da luz.

A lição filosófica a ser tirada da classificação acima é – atenção – a *condição física da causalidade*: se dois eventos X e Y estão ligados por um intervalo do tipo espaço, então não pode haver relação causal entre eles (dizemos X e Y são “causalmente desconectados”); se o intervalo a conectar X e Y for do tipo tempo, então, X pode influenciar Y (ou o inverso) causalmente através de uma partícula ou evento material que tenha velocidade inferior a c ; se o tipo de intervalo a conectar X e Y for do tipo luz, então, a influência de um evento no outro, se existir, deve se propagar na velocidade da luz.

A figura 4 – que denota o famoso “cone de luz” da relatividade restrita – codifica os fatos acima descritos sobre a causalidade. Incentivamos o estudante a gastar algum tempo interpretando esta figura. Sugestão: para simplificar, interprete a figura 4 imaginando um espaço unidimensional; identifique a reta do intervalo tipo luz no desenho com a fórmula apresentada na definição A acima e com a fórmula de reta que você deve ter aprendido na escola; em seguida,

identifique as inequações das definições B e C na figura.

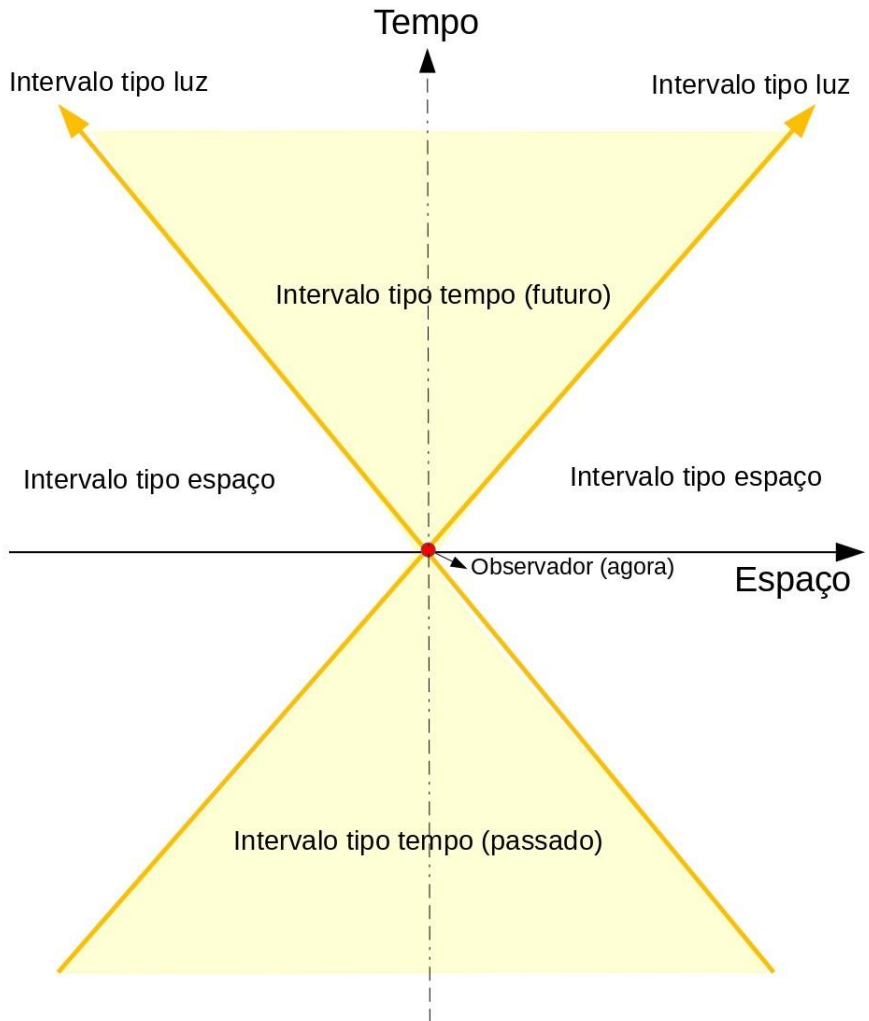


Figura 4. O cone de luz e a estrutura causal do espaço-tempo.

Agora que entendemos algumas noções básicas da relatividade restrita, podemos partir para estudar como se dá a transformação de coordenadas na teoria. Para começar, apresentamos ao estudante o que chamamos, em física, “relógio de luz”. Considere, em um referencial inercial S, dois espelhos planos dispostos paralelamente ao eixo x, com faces refletoras uma de frente para a outra a uma distância L_0 fixa. Um sinal luminoso mede o tempo ao registrar um “tic” cada vez que a luz bate no espelho de baixo:

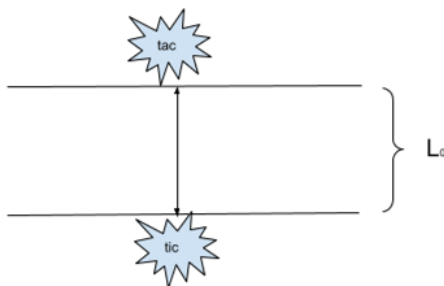


Figura 5. O espelho de luz no referencial S.

Recorde: “período” representa o tempo necessário para que um movimento volte a se repetir. Aplicando a definição de velocidade, portanto, o período do relógio de luz é

$$T_0 = \frac{2L_0}{c}. \quad (5)$$

Questão: como alguém que está em um referencial S' em movimento relativo constante em relação a S enxerga o relógio de luz?

Lembre-se do princípio de relatividade: se para o observador em S (Kant, digamos), S' se move com velocidade constante em linha reta para a esquerda, para o observador em S' (Hegel, digamos), S se move com velocidade constante para a direita. Pelo postulado de Einstein de que as leis da física devem ser as mesmas em todos os referenciais, Hegel e Kant devem observar o mesmo período para o relógio de luz. Mas, como para Kant o relógio de luz está

se movendo para a direita (eixo x) com velocidade constante v , Hegel, em S' , vê o tic-tac do relógio em S como na figura 6.

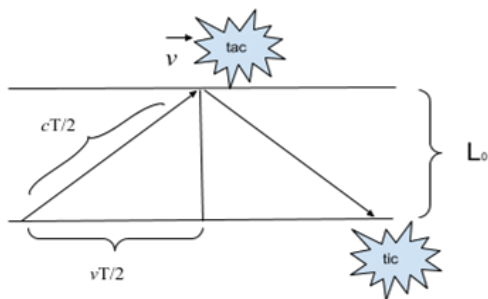


Figura 6. O espelho de luz no referencial S' .

Sendo T o intervalo entre os eventos “tic” e “tac” do relógio de luz em S , pelo segundo postulado de Einstein e pelo Teorema de Pitágoras, portanto,

$$\left(\frac{cT}{2}\right)^2 = (L_0)^2 + \left(\frac{vT}{2}\right)^2 \quad (6)$$

e assim, substituindo (5) em (6) para eliminar L_0 (faça-o!), temos

$$T = \frac{T_0}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} = \gamma T_0, \quad (7)$$

onde γ – o chamado “fator de Lorentz” – vale $\sqrt{1 - \left(\frac{v}{c}\right)^2}^{-1}$. Lembre que v é sempre menor que c ; portanto, a quantidade γ é sempre maior ou igual a 1. Assim, $T = \gamma T_0$ nos diz:

para S' o relógio se move mais devagar do que para S .

A fórmula (7) reflete o que chamamos, em relatividade restrita, “dilatação do tempo”.

Conclusão: em um referencial inercial S' que se move com velocidade constante e em linha reta em relação a um referencial inercial em repouso S o tempo passa mais devagar.

Agora, note o seguinte. Imagine que Kant tenha não apenas um, mas dois relógios de luz em S : um deles com os espelhos dispostos paralelamente ao eixo x (figura 5) e outro, idêntico ao primeiro, mas com os espelhos dispostos paralelamente ao eixo y . Ora: S é um referencial inercial. Logo, ambos os relógios se movem com o mesmo período T_0 .

Seja L a distância entre os espelhos do segundo relógio de luz, considere o período de tempo T que a luz leva para ir e voltar observado por Hegel em S' : começando pelo espelho de trás (que corresponde ao espelho de baixo na figura 5) a luz atinge o espelho da frente (tic) após um tempo t dado por $ct = L + vt$ e retorna (tac) em um tempo t' dado por $ct' = L - vt'$. Ora! $T = t + t'$. Portanto,

$$T = \frac{2L}{c(1-\frac{v^2}{c^2})} \quad (8)$$

e, como os relógios devem ter o mesmo período, a partir de (7) e (8) temos

$$\frac{2L}{c(1-\frac{v^2}{c^2})} = \gamma T_0. \quad (9)$$

Por fim, substituindo a expressão (5) para T_0 em (9) (faça-o!) vemos que, para Hegel, a distância entre os espelhos do segundo relógio de Kant em S é

$$L = L_0 \sqrt{1 - \frac{v^2}{c^2}} = \frac{L_0}{\gamma}. \quad (10)$$

Repetimos: γ é sempre maior ou igual a 1. Logo, de $L = \frac{L_0}{\gamma}$ descoberta:

para S' a distância entre os espelhos paralelos ao eixo y é menor do que para S .

A fórmula (10) reflete o que chamamos, em relatividade restrita, de “contração do espaço”.

Conclusão: se um referencial inercial S' se move com velocidade constante e em linha reta em relação a um referencial inercial em repouso S , e se um objeto, em S , está disposto paralelamente à direção do movimento, então, os observadores em S e S' – Kant e Hegel, no nosso exemplo – *discordarão sobre o tamanho do objeto*.

Em posse desses resultados, derivaremos, em seguida, as transformações de Lorentz para translações no eixo x .

Considere novamente os referenciais inerciais S e S' e sejam (x, y, z, t) e (x', y', z', t') as coordenadas de um evento Ω em S e S' respectivamente. S' se move em relação a S com velocidade constante v na direção do eixo x ; O e O' denotam as *origens espaciais* dos referenciais S e S' ; O e O' coincidem em $t = t' = 0$.

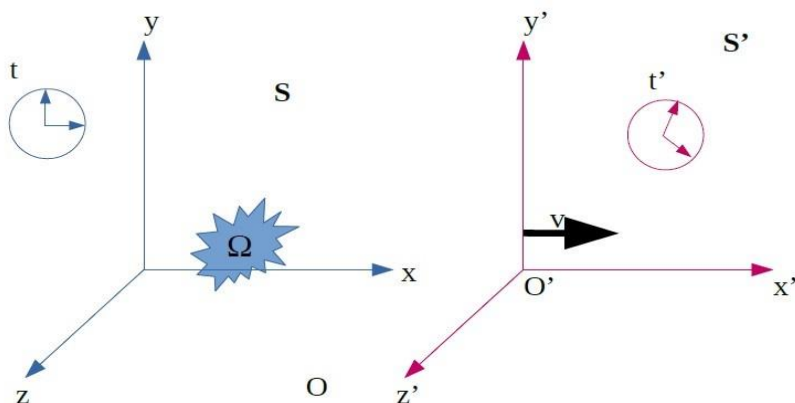


Figura 7. Referenciais inerciais S (em azul) e S' (em vermelho). S' está em movimento em relação a S com velocidade constante v na direção do eixo x . O e O' coincidem em $t = t' = 0$.

Para derivar as transformações de Lorentz para translações, é suficiente que utilizemos os resultados anteriores sobre contração do espaço e dilatação do tempo, e a relação

$$\overline{O\Omega} = \overline{OO'} + \overline{O'\Omega} \quad (11)$$

entre vetores (que aqui se reduzem a segmentos orientados), nos referenciais S e S'. Definindo explicitamente esses vetores no referencial S:

$$\begin{aligned} \overline{O\Omega} &= x \text{ (distância euclidiana entre } \Omega \text{ e } O\text{),} \\ \overline{OO'} &= vt \text{ (distância euclidiana entre } O \text{ e } O'\text{),} \\ \overline{O'\Omega} &= \frac{x'}{\gamma} \text{ (distância euclidiana entre } O' \text{ e } \Omega\text{) pois, atenção, } x' \text{ é } \overline{O'\Omega} \text{ para } S' \text{ (vide equação (10)).} \end{aligned}$$

Substituindo isso em (11), temos

$$x = vt + \frac{x'}{\gamma} \quad (12)$$

ao passo em que, para S',

$$\begin{aligned} \overline{O\Omega} &= \frac{x}{\gamma} \text{ (distância euclidiana entre } \Omega \text{ e } O\text{) pois – atenção novamente – } x \text{ é } \overline{O\Omega} \text{ para } S \text{ (vide novamente equação (10)),} \\ \overline{OO'} &= vt' \text{ (distância euclidiana entre } O \text{ e } O'\text{),} \\ \overline{O'\Omega} &= x' \text{ (distância euclidiana entre } O' \text{ e } \Omega\text{),} \end{aligned}$$

portanto, por (11),

$$\frac{x}{\gamma} = vt' + x'. \quad (13)$$

Rearranjando termos em (12) (faça-o!), obtemos, portanto,

$$x' = \gamma(x - vt) \quad (14)$$

que é a transformada de Lorentz entre dois referenciais S e S' em movimento relativo na direção x. Sua inversa – lembre-se: se S' se move com velocidade constante e em linha reta para a direita em relação a S, para S', S se movimenta com velocidade constante e em linha reta para a esquerda – é obtida ao se reorganizar (13) (faça-o!):

$$x = \gamma(x' + vt').(15)$$

A transformada de Lorentz temporal é obtida substituindo-se (14) em (15) (faça-o!):

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)(16).$$

Note quão curioso: t' em (16) depende da coordenada x e da velocidade relativa v!

As transformações de Lorentz entre S e S' (figura 7) estão esquematizadas na tabela abaixo. O estudante deve compará-las com as transformações de Galilei indicadas no início desta seção (em particular no que diz respeito à coordenada temporal); as coordenadas y e z não mudam porque são perpendiculares à direção do movimento.

Tabela 1. Transformadas de Lorentz e suas inversas para sistemas de referência S e S' em movimento relativo v. Coordenadas coincidem em um dado instante referido como “instante zero” em ambos os referenciais.

TL entre S e S' (configuração padrão)	TL inversa entre S e S' (configuração padrão)
$x' = \gamma(x - vt)$	$x = \gamma(x' + vt')$
$y' = y$	$y = y'$
$z' = z$	$z = z'$
$t' = \gamma\left(t - \frac{vx}{c^2}\right)$	$t = \gamma\left(t' + \frac{vx'}{c^2}\right)$

Corolário. Se você perguntar a um observador galileano o que significa dois eventos serem simultâneos, a resposta será: “os eventos acontecem ao mesmo tempo”. Isso significa que eventos $X = (x_1, t_1)$ e $Y = (x_2, t_2)$ são simultâneos para um observador galileano se e somente se $t_1 = t_2$. Como, lembre-se, $t = t'$ para observadores galileanos em referenciais inerciais S e S' (vide página 369), se X e Y simultâneos para S , então, X e Y simultâneos para S' . O estudante deve observar que essa noção de simultaneidade deve ser abandonada à luz dos resultados acima apresentados e discutidos. Em referenciais lorentzianos, $t \neq t'$. Logo, *observadores lorentzianos podem discordar sobre a ordem de eventos*. A depender das coordenadas e da velocidade envolvidas, se X e Y simultâneos em S , X pode ocorrer antes ou depois de Y em S' (ou vice-versa).

A simultaneidade não é um absoluto.

A proposta deste texto introdutório não comporta discutir as interessantíssimas representações gráficas destes fatos, mas incentivamos o leitor a procurar por elas; sugerimos a busca por conceitos como *linha de mundo*, *hiperplano*, *planos de simultaneidade*, e *espaço de Minkowski*.

Agora que entendemos alguns conceitos fundamentais da teoria da relatividade restrita, exploremos algumas de suas consequências mais simples e intrigantes. A seguinte situação, descrita por Paul Langevin – renomado físico francês – na obra “L'évolution de l'espace et du temps” ilustra o quão drasticamente o tempo é afetado pelo movimento:

Bastaria para isso que nosso viajante consentisse de se fechar dentro de um projétil que a Terra lançaria com uma velocidade suficientemente próxima daquela da luz, porém inferior, o que é fisicamente possível, arranjando-se para que se produzisse um encontro, com uma estrela por exemplo, após um ano da vida do viajante, e lhe mandasse para a Terra com a mesma velocidade.

De volta à Terra, tendo envelhecido dois anos, ele sairá de sua arca e encontrará nosso globo envelhecido em duzentos anos, se sua velocidade permaneceu dentro de um intervalo de somente um vigésimo milésimo a menos que a velocidade da luz. Os fatos experimentais mais seguramente estabelecidos da física nos permitem afirmar que assim seriam as coisas. (p. 50)

Analisemos a situação descrita por Langevin mais cuidadosamente. Imagine que duas pessoas, Pedro e Paulo – irmãos gêmeos idênticos – tenham decidido levar a cabo o experimento mental proposto por Langevin. Pedro fica na Terra, enquanto Paulo se fecha em um projétil, viaja até uma estrela a 99.995 anos-luz distante da Terra, e volta em seguida.

De acordo com Langevin, a velocidade de Paulo é

$$v_{\text{Paulo}} = \left(c - \frac{c}{20.000} \right) = \left(1 - \frac{1}{20.000} \right) c = 0.99995c \quad (17)$$

e portanto, *para Pedro*, no referencial da Terra,

$$\text{tempo de viagem de Paulo} = \frac{2 \times (\text{distância})}{\text{velocidade de Paulo}} = \frac{199.99 \text{ anos-luz}}{0.99995c} = 200 \text{ anos.} \quad (18)$$

Sendo $v = 0.99995c$, $\gamma = 100$ (faça, você, também as contas!). Logo, por (7),

$$T_{(\text{viagem de Paulo no referencial de Paulo})} = \frac{T_{(\text{viagem de Paulo no referencial de Pedro})}}{\gamma} = 2 \text{ anos.} \quad (19)$$

Pause a leitura para sentir o impacto metafísico desse resultado, derivado com alguma facilidade, sobre as noções de espaço e tempo newtonianas. Perceba como ele se contrasta com sua experiência cotidiana de espaço e de tempo. Note, contudo, que isso não significa que as posições de Leibniz estavam corretas. Segundo Einstein, o espaço e o tempo – ou, deveríamos dizer, o espaço-tempo – são conceitos mais complexos do que Newton, Leibniz e muitos outros pensadores supunham ser. Ademais, as consequências dos postulados de Einstein mostram com clareza: não é possível

investigar sobre a natureza da realidade *a priori*, “da poltrona do filósofo”, como se costuma dizer. A investigação filosófica e científica sobre a natureza do mundo são interdependentes.

Há um outro aspecto nesse resultado sobre a viagem de Paulo que causa perplexidade. Sabemos que as leis da física são as mesmas em referenciais inerciais. Se para Pedro, Paulo se afasta em uma determinada direção com velocidade constante v , para Paulo, Pedro é que se afasta com velocidade constante $-v$. Mas, se assim é – se a situação física é simétrica, isso é dizer – como pode Paulo viajar a uma estrela e, ao voltar, não encontra mais seu irmão?

Com estas considerações, apresentamos ao estudante o famoso “paradoxo dos gêmeos” da relatividade restrita. Note o quão expressivamente t depende de v : se $v_{\text{Paulo}} = 0.98c$, $\gamma = 5.025$, e portanto, se Paulo faz uma viagem de ida e volta de 2 anos, para Pedro, passam-se “apenas” 10 anos – o “paradoxo” está em um dos gêmeos ter envelhecido, biologicamente, 8 anos a mais que o outro.

Você talvez esteja se perguntando: há, então, uma inconsistência na teoria de Einstein? Tranquelize-se: o “paradoxo” é apenas aparente (o que justifica nosso uso das aspas). Atente-se à situação física que Langevin descreve e, antes de prosseguir a leitura, aplique aí os postulados da teoria, procurando identificar o erro de raciocínio que leva ao “paradoxo”.

Verifique se sua constatação coincide com o que segue: Paulo, ao manobrar o foguete para retornar à Terra, *não estava mais em movimento relativo uniforme, mas em um movimento acelerado*. Isso significa que a simetria inicial entre os referenciais foi quebrada. Paulo passou de um referencial (ida) para outro (volta), enquanto Pedro manteve-se no mesmo referencial desde o início da viagem de Paulo.

Observe que Paulo – em um certo sentido estrito – “viaja para o futuro”: como mostrado nas expressões (18) e (19), Paulo diminuiu a velocidade do relógio que marca o tempo no seu referencial (isso inclui seu relógio biológico, é claro!) relativamente ao relógio de Pedro, de tal modo que o período do tempo no referencial dele, e o período do tempo no referencial da Terra, *estão fora de fase*. Porém, é de grande importância ter em mente que essa “viagem no tempo” é *uma viagem só de ida*: qualquer viagem que Paulo faça – a velocidades relativísticas, é claro – irá lançá-lo *cada vez mais adiante no futuro da Terra*.

Antes de encerrarmos essa elementar introdução à noção de espaço-tempo, vamos apresentar mais uma relevante consequência dos postulados de Einstein. Esta consequência é particularmente interessante, pois diz respeito à natureza do tempo, assunto que é caro aos filósofos desde a Antiguidade. O argumento a ser apresentado deve-se a Hilary Putnam (1967). Em resumo, ele afirma que a relatividade da simultaneidade, que estudamos acima, implica, sob algumas hipóteses adicionais, em uma tese metafísica radical e profunda sobre a natureza da realidade: o chamado “eternalismo”, ou, “a tese do universo fechado”. Mais precisamente, essa tese deriva de pensar *a realidade* como uma *relação de equivalência* e, em seguida, aplicar os resultados da relatividade a essa realidade. Sendo assim, comecemos pela definição do abstrato conceito de relação de equivalência:

Seja X um conjunto qualquer não-vazio, dizemos que uma relação $R \subset X \times X$ é uma “relação de equivalência em X ” se R satisfizer os seguintes quesitos:

1. *Reflexividade*. $\forall a \in X, (a, a) \in R$.
2. *Simetria*. Se $(a, b) \in R, (b, a) \in R$.
3. *Transitividade*. Se $(a, b) \in R, (b, c) \in R$, então $(a, c) \in R$.

Caso o leitor tenha dificuldade com a notação utilizada na definição, não se preocupe; isso é natural, especialmente se o leitor for um estudante de graduação calouro. Se necessário, pause a leitura para digerir as informações e/ou para pesquisar o significado dos símbolos utilizados.

Agora, apliquemos a definição. Assumamos que “realidade” denota o conjunto de coisas que são simultâneas para um observador em cada instante; entendamos X como “universo” (no sentido de “a totalidade das coisas que existem”) e R como “realidade” (tenha em mente que R é subconjunto de $X!$); note que se segue daí, através do quesito de transitividade, que, se eu e você somos reais, e algo é real para você, então, esse algo também é real para mim. Portanto, retomando a narrativa dos personagens Pedro e Paulo:

1. O “agora” de Pedro é real (isso vale para todo e cada instante t , é claro).
2. Há, pelo menos, um outro observador – Paulo, no nosso exemplo – que é real, e pode, ademais, estar em movimento relativamente a Pedro.
3. Se todas as coisas que estão em relação R com o agora de Pedro são reais, e se o agora de Paulo é real, então, todas as coisas que estão em relação R com o agora de Paulo estão em relação R com o agora de Pedro.

Em outras palavras, ainda que certos eventos que estão no agora-de-Paulo façam parte apenas do futuro-de-Pedro, a realidade dos eventos no agora-de-Paulo vale também para Pedro; a realidade não pode depender do referencial em que o observador está. Daí a descoberta:

O futuro já existe.

Note que, na física clássica, o tempo – absoluto para todos os observadores – é distintivamente dividido em “passado, presente e futuro”. A teoria da relatividade, porém, nos leva a uma situação mais complicada: eventos que são “simultaneamente” passado-e-futuro (passado para Paulo, futuro para Pedro); eventos que estão fora do limite de causalidade para Pedro, mas não estão para Paulo (o intervalo entre esses eventos, isto é dizer, são do tipo espaço para Pedro, do tipo tempo para Paulo).

Para fortalecer o argumento, vamos rephrasing a conclusão extraída das premissas 1-3 sobre o futuro: a relatividade da simultaneidade nos diz que certos eventos no futuro de Pedro já aconteceram ou estão acontecendo para Paulo; se Pedro e Paulo são parte da realidade, e se os eventos presentes para Paulo são parte da realidade, então, *embora tais eventos ainda não tenham acontecido para Pedro, eles existem*. Assim, a noção newtoniana que trata o tempo como uma variável na qual pontos são continuamente adicionados – a noção dinâmica de que a cada novo instante há um novo universo, e de que o universo do instante passado deixa de existir – deve ser abandonada. Conclui-se do argumento de Putnam, portanto, que

A realidade é um bloco fechado.

Você pode pensar do seguinte modo: assim como você possui “partes espaciais” – você existe em certas partes do espaço e não em outras, mas o espaço existe “todo de uma vez” –, da mesma forma você possui “partes temporais”: você existe – de diferentes maneiras – em diferentes “partes” do tempo, mas o tempo, como o espaço, existe “todo de uma vez”:

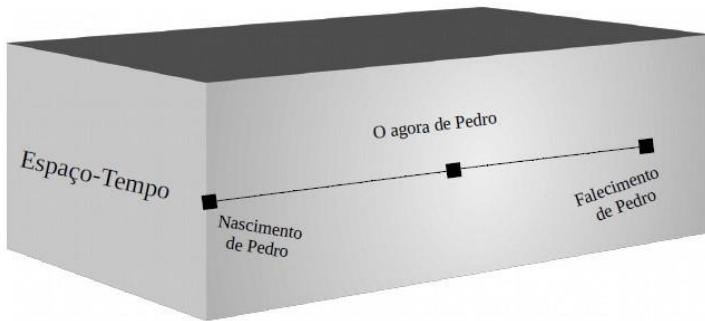


Figura 8. O universo em bloco. Todos os eventos do bloco – entendido como uma entidade quadridimensional – têm o mesmo status ontológico. Seu futuro e seu passado, isso é dizer, não são diferentes do seu agora. Não há “temporalidade” associada à sua existência.

Não nos aprofundaremos neste debate. Há sutilezas a serem consideradas no emprego das premissas 1-3, e é possível – com argumentos filosóficos sofisticados – chegar a conclusões diferentes. Ademais, ingredientes da teoria da relatividade geral e de teorias de gravidade quântica sugerem argumentos adicionais, mais complexos, remodelando o debate. Nos damos por satisfeitos se essa apresentação dos resultados de Hilary Putnam servir para que o leitor perceba como uma teoria da física pode ter consequências sobre grandes questões metafísicas, e que, portanto, os argumentos da ciência podem ser poderosas armas para o filósofo. Esse é mais um exemplo da opacidade da

fronteira entre filosofia e física.

Aqui encerramos a seção sobre espaço, tempo e movimento. Vamos enumerar e resumir as mais importantes conclusões que obtivemos: (1) segundo a proposta de Einstein, a noção newtoniana de tempo e espaço absolutos estão erradas; isso, contudo, não equivale a dizer que a posição de Leibniz, antagônica à de Newton, equivale à de Einstein. A teoria da relatividade restrita leva a conclusões muito mais amplas e complexas comparadas às ideias de Leibniz; (2) os conceitos de intervalo e cone de luz são exemplos do importante papel da relatividade em discussões sobre causalidade. Não estamos afirmando que a relatividade coloca um ponto final nesse debate, mas, sim, que estabelece uma arena onde essas discussões devem ocorrer; (3) vimos como a relatividade restrita altera as noções clássicas de ordenação de eventos e de simultaneidade. Consequentemente, a teoria tem forte influência no debate acerca da noção de realidade, debate este que, talvez, o leitor pensava ser externo à ciência.

A mensagem principal desta seção, contudo, não diz respeito às teorias de Newton, Leibniz e Einstein. Não temos a pretensão de acreditar que somos capazes de ensinar todas as importantes questões relativas a essas teorias em tão poucas páginas. Almejamos, porém, que o estudante leve deste texto a compreensão da interdependência entre filosofia e física (e todas as demais ciências, é claro). A física não só se alimenta de metafísica, mas também produz metafísica. A natureza da realidade depende das condições físicas do mundo. Não é possível que o filósofo as descubra do assento de sua poltrona. Tampouco é possível que o cientista o faça sem que recorra a modos de pensar filosóficos. Fazer ciência é uma aventura filosófica. O estudante talvez sinta o peso deste fato ainda maior força na seção seguinte, onde tratamos de difíceis questões, muitas das quais ainda em aberto, ligadas à teoria quântica.

2. Teoria quântica: física, filosofia, problemas

Abrimos esta seção com algumas advertências. Em primeiro lugar, a teoria quântica, hoje, é provavelmente a mais bem-sucedida teoria de toda a física: seu sucesso, neste contexto, é atribuído ao fato de toda a tecnologia que transformou a humanidade ao longo do século XX ser fundamentada na teoria quântica. Três das quatro forças da Natureza – a eletromagnética, a nuclear fraca e a nuclear forte – são descritas por teorias quânticas, a saber, a

eletrodinâmica e a cromodinâmica quânticas. Essas teorias compõem o chamado “Modelo Padrão de Partículas Elementares”, também conhecido como “o campeão teórico invicto da precisão experimental”: a concordância entre previsões teóricas e resultados experimentais chega, no caso da eletrodinâmica quântica, a onze casas após a vírgula. O estudante deve avaliar esse fato tendo em mente as lições da filosofia da ciência: a verificação de hipóteses científicas é feita por experimentos. Nenhum experimento, até hoje, falseou a teoria quântica. Logo, o mundo – para mencionar o mote dos físicos – é quântico. Logo, a física clássica é falsa.

Talvez o peso dessa afirmação perturbe você. A física quântica – em particular, a mecânica quântica – surgiu como reação dos físicos à realização de que as previsões da física clássica, no caso da interação entre radiação e matéria, estarem em flagrante contradição com os fatos. Você estuda física clássica na escola, a física clássica é parte da física. Porém, você já deve ter concluído da seção anterior que o mundo não é como Newton o concebeu, o mundo é *relativístico*. Por que, então, estudar física clássica? Atente-se às equações (15)-(16): para $v \ll c$, ≈ 1 , as transformações de Lorentz reduzem-se às transformações de Galilei. O fato é: *no limite de baixas velocidades* – o chamado “limite não-relativístico” – as equações da física clássica são *válidas* – não verdadeiras! – “para todos os fins práticos”. A atitude dos cientistas ante a física clássica é pragmática; do ponto de vista metafísico, contudo, o mundo é relativístico, não clássico. E é também quântico. As teorias acima mencionadas, eletrodinâmica e cromodinâmica quânticas, são teorias quânticas relativísticas (teorias quânticas *de campos*). E da mesma maneira com que “ignoramos a relatividade” no limite de baixas velocidades, “podemos ignorar a teoria quântica” na escala supra-atômica, pois aí, os efeitos quânticos não se manifestam. Mas, repetimos, fundamentalmente, o mundo é quântico. Este é um fato experimental bem estabelecido. As leis que regem a Natureza são determinadas por princípios quânticos; tal fato subjaz a maior parte da física estudada e desenvolvida nos séculos XX e XXI.

O problema é que não sabemos bem quais são ou o que dizem esses princípios. Não entendemos boa parte do que a teoria quântica diz sobre a realidade. A segunda advertência que fazemos ao estudante diz respeito à atitude dos físicos sobre esse estado de coisas. O físico médio, em seu cotidiano, não lida com essas questões. É claro que quando um cientista trabalha em

desenvolvimento de tecnologia, por exemplo, problemas ligados aos fundamentos da física são pouco relevantes. Em outros casos, porém, a situação é mais grave; alguns físicos se opõem à reflexão sobre os problemas da mecânica quântica. Estudantes interessados são com frequência desencorajados; parece existir uma regra tácita na prática de muitos físicos, segundo a qual o fazer perguntas a respeito do significado da teoria é “perda de tempo”, pois afinal, “o formalismo funciona muito bem”: a atitude popular é resumida pelo físico David Mermin nos termos “cale a boca e calcule”.

A atitude anti-filosófica não caracteriza, porém, a prática dos autores da mecânica quântica. Max Planck, Albert Einstein, Niels Bohr, Werner Heisenberg, Louis de Broglie, Erwin Schrödinger, Max Born desenvolveram a teoria, tendo as questões conceituais como eixo e ponto de partida. O instrumentalismo e cientificismo anti-filosóficos são um fenômeno sociológico que se solidificou no pós-guerra (vide Kragh 2010 para detalhes). Fazemos notar, porém, e enfatizamos, que a teoria quântica coloca problemas sobre a natureza da realidade, sobre a objetividade da experiência, sobre a relação ontológica entre um sistema complexo e suas partes, sobre a causalidade, entre muitos outros, que são filosóficos em essência. Muitas das questões sobre os fundamentos da teoria que tiravam o sono de N. Bohr e A. Einstein acompanham-nos até hoje. A situação, “ninguém entende a teoria quântica”, como famosamente pronunciou Richard Feynman, é inaceitável porque a teoria é uma teoria física, seu objeto é a realidade – sendo teorias criações do intelecto humano, sendo a realidade inteligível, por certo a teoria se deve fazer inteligível. A última advertência que fazemos ao estudante diz respeito ao fato, já enunciado, desses problemas estarem em aberto. Repetimos que eles são problemas filosóficos, adicionamos o fato de a natureza deles ser tal que é impossível entendê-los propriamente, quanto menos atacá-los, sem dominar o formalismo técnico. Não é possível ir além dos limites da teoria quântica, se você não conhecer seus limites. Você deve, portanto, para bem realizar sua tarefa filosófica, ter pleno domínio do que a teoria é capaz de fazer.

A apresentação feita nesta seção é extremamente simplificada e serve apenas para dar ao estudante uma noção rudimentar dos problemas filosóficos que existem em fundamentos de mecânica quântica não-relativística. Questões semelhantes aparecem, em nível de dificuldade muito maior, em teoria quântica de campos e em teorias de gravidade quântica. Não trataremos dessas questões

aqui. Vale, porém, observar que extensões relativísticas da mecânica quântica herdam os problemas do caso não-relativístico.

2.1. Dualidade onda-partícula

No primeiro parágrafo do primeiro livro da obra *Metafísica*, Aristóteles escreve: “a visão nos proporciona mais conhecimentos do que todas as outras sensações, e nos torna manifestas numerosas diferenças entre as coisas”. Ora, a percepção da visão é circunstanciada pela presença da luz – mas o que, afinal de contas, é a luz?

Vejamos. A teoria corpuscular da luz – devida a Descartes e Newton – há tempos houvera sido relegada ao esquecimento: a luz era conhecida por ser uma onda; mais precisamente, uma solução periódica das equações do eletromagnetismo de Maxwell. O experimento da dupla fenda, proposto por Thomas Young em 1802, foi considerado o “experimento crucial” que demonstrou – conclusivamente, para os físicos do século XIX – a natureza ondulatória da luz:

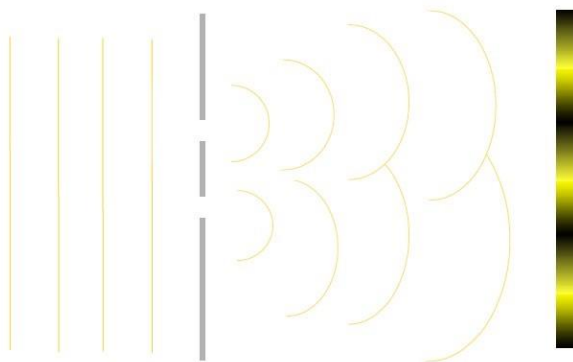


Figura 9. O experimento da dupla fenda de Young. Um feixe luminoso é lançado em direção a um anteparo onde há duas fendas e detectado, em seguida, em uma tela situada a certa distância das fendas. As faixas claras e escuras indicam as regiões de interferência construtiva e destrutiva, respectivamente. Partículas clássicas não interferem nem construtiva nem destrutivamente; portanto, a luz é uma onda.

Na seção anterior vimos que Einstein, em 1905, chegou à teoria da relatividade restrita levando metafisicamente a sério as leis do eletromagnetismo. Ainda assim, o próprio Einstein, meses antes de publicar o artigo sobre a relatividade, ousou contestar os resultados do eletromagnetismo:

não obstante a verificação experimental completa da teoria da reflexão, refração, dispersão, etc., é perfeitamente concebível que uma teoria da luz envolvendo funções contínuas no espaço levará a contradições à experiência quando aplicada aos fenômenos de criação e conversão de luz... de fato, as observações sobre a radiação de corpo negro, fotoluminescência, produção de raios catódicos [...] e outros fenômenos envolvendo a emissão e conversão de luz podem ser melhor compreendidos na suposição de que a energia é distribuída *descontinuamente* no espaço [...] e consiste em um *número finito de quanta de energia, localizado no espaço*, que se move *sem ser dividido* e que só pode ser absorvido ou emitido *como um todo*. (1905a, p. 133)

Nas palavras de Francis Bacon, “só se pode dominar a natureza obedecendo-a”. E o fato é, contra o esperado em vista da física clássica, que o fenômeno do efeito fotoelétrico só pode ser explicado, caso admitamos, como Einstein sugere, que a luz é composta por “pacotinhos de energia”, isto é, em suma, partículas, que Einstein denominou “quantas”. Explicamos esse fato no que segue.

O experimento de efeito fotoelétrico consiste, resumidamente, em expor uma peça metálica à radiação eletromagnética e detectar a corrente elétrica que a peça produz como resultado dessa exposição. Mais precisamente, quando radiação eletromagnética incide sobre um metal, elétrons podem absorvê-la; se a energia transferida da radiação para o elétron for suficientemente alta, o elétron pode ser ejetado do material e detectado em forma de corrente elétrica. Façamos uma analogia com uma situação com a qual o estudante deve estar mais familiarizado: imagine um corpo repousando sob uma prateleira como na figura 10. O lado esquerdo da prateleira é fechado e o lado direito é aberto. Por

simplicidade, imagine que o corpo está preso à prateleira como um trem fica preso ao trilho, de modo que ele só pode se mover ao longo da prateleira. Sendo assim, para remover o corpo da prateleira, é necessário movê-lo até o limite direito desta. A situação pode ser descrita, como os físicos costumam gostar, em termos de energia: para que o corpo possa sair da prateleira, é necessário dar-lhe energia até que ele vença a barreira de potencial representada por V ; em outras palavras, dizer que o corpo precisa alcançar uma certa altura é o mesmo que dizer que sua energia potencial gravitacional deve alcançar um dado valor. Algo muito semelhante acontece com o elétron preso ao sólido metálico, isto é, o elétron sente forças de atração dos átomos que compõem o metal. Assim como a força da gravidade submete o corpo na prateleira a um potencial, as forças de atração submetem o elétron a um potencial que precisa ser vencido caso se pretenda libertar o elétron.

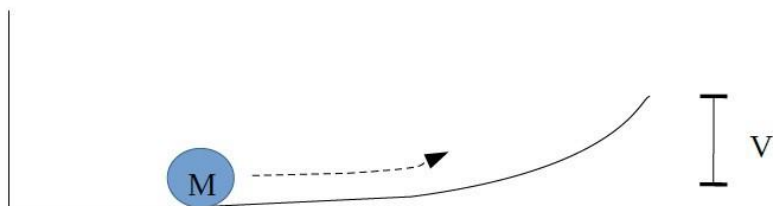


Figura 10. Bola em prateleira com barreira de potencial V .

Segundo o eletromagnetismo clássico, não há restrições quanto às características da radiação, para que o elétron possa ser ejetado do metal. Se a radiação tiver baixa intensidade – em outras palavras, se ela transfere pouca energia por segundo para os elétrons –, os elétrons podem absorver a energia eletromagnética continuamente por algum tempo até serem ejetados, ou seja, basta aguardar tempo suficiente para observar os elétrons saindo do metal. Na analogia da prateleira, isso seria equivalente a dizer que o corpo pode receber energia continuamente sendo empurrado aos poucos prateleira a cima até vencer o potencial V .

A experiência, contudo, mostra resultados muito distintos desses. O que se observa, ao incidir radiação eletromagnética sobre um metal, é que o fator determinante, no que diz respeito à capacidade de emitir ou não elétrons, é

a *frequência*⁹⁷ da radiação. Mais precisamente, radiação com frequência igual a ou maior do que um determinado valor é capaz de ejetar os elétrons; radiação com frequência menor do que esse valor não é capaz de ejetar elétrons. Curiosamente, a intensidade da radiação é irrelevante quanto à possibilidade de emissão dos elétrons. Por exemplo, a depender do metal, quando um feixe de baixa frequência – por exemplo, de cor vermelha – é utilizado, ainda que esse feixe seja muito intenso, nenhum fluxo de elétrons saindo do metal é detectado; por outro lado, quando um feixe de alta frequência – por exemplo, um feixe ultravioleta – é utilizado, ainda que ele seja pouco intenso, um fluxo de elétrons ejetados é observado.

Einstein propôs uma solução muito simples para o problema. Se o elétron precisa de uma quantidade mínima de energia (essa energia é geralmente chamada de *função trabalho*, ϕ) para vencer o potencial e tornar-se livre, e existe uma frequência mínima experimentalmente observada que é necessária para a emissão dos elétrons, então, concluiu Einstein, deve haver uma relação direta entre a energia que os elétrons recebem e a frequência da radiação. Daí vem a hipótese dos quanta; isto é, Einstein propôs que a radiação deve ser composta por unidades elementares que contêm uma certa quantidade de energia (da mesma maneira que pensamos que a matéria é composta por unidades elementares de massa chamadas de átomos); e que essa energia E deve ser proporcional à frequência ν da radiação de acordo com a fórmula $E = h\nu$, onde h é a constante de Planck (em SI, $6,62607004 \times 10^{-34}$ m² kg/s). Essas unidades elementares de energia são chamadas *fótons*.

Assim, um feixe de radiação de baixa frequência é formado por numerosos fótons com baixa energia; a diferença deste feixe para um outro mais

⁹⁷Aqui definimos este conceito para os leitores que não estejam familiarizados com ele. Frequência é a quantidade de oscilações por unidade de tempo. Uma onda que oscila “muito rapidamente” tem “alta frequência”, uma onda que oscila “devagar”, “baixa frequência”. O ser humano percebe a frequência da radiação eletromagnética como cor. A cor vermelha tem a menor frequência que o ser humano consegue enxergar. A cor violeta tem a maior. Daí vem as nomenclaturas “radiação infravermelha” (radiação que tem frequência menor que a da cor vermelha, isto é, menor que nosso limite inferior de detecção, e que, portanto, não enxergamos) e “radiação ultravioleta” (radiação que tem frequência maior que a da cor violeta, isto é, maior que nosso limite superior de detecção, e que, portanto, não enxergamos).

intenso e de mesma frequência é simplesmente a quantidade de fótons que o compõem. Isso explica o porquê de os elétrons não serem ejetados a partir de feixes vermelhos no exemplo citado: os fótons que compõem o feixe vermelho têm energia menor que o mínimo necessário para libertar os elétrons. Não interessa quão intenso seja o feixe vermelho, seus fótons não têm a energia necessária para libertar os elétrons. Por outro lado, não interessa quão fraco seja o feixe ultravioleta, seus fótons sempre terão energia suficiente para ejetar os elétrons do metal. Colocando em outros termos:

A radiação eletromagnética, que pensávamos se comportar estritamente como onda, aparenta ser formada por partículas (fótons); a interação da luz com a matéria pode ser descrita de modo semelhante a que usamos para descrever a colisão entre bolas de bilhar.

Para encerrar a ilustração da prateleira, a troca de energia entre a radiação e os elétrons é uma situação análoga a tentar libertar o corpo da prateleira arremessando nele uma bola de bilhar. Se a bola de bilhar (fóton), ao colidir com o corpo (elétron), tem velocidade (frequência) suficientemente alta, ela poderá transmitir energia suficiente para libertar o corpo (elétron) da prateleira (do metal).

Atento a este resultado, considere, agora, a seguinte pergunta: se a luz, que supúnhamos, motivados por experimentos, ser uma onda, se comporta sob determinadas condições experimentais como uma partícula, reciprocamente também poderia o elétron – que sabemos, por experimentos, ser uma partícula – se comportar, sob condições experimentais específicas, como onda?

Vejamos. Considere um arranjo experimental como o da figura 9. Em vez de um raio luminoso, porém, imagine que *elétrons* sejam lançados – um por vez – contra o anteparo no qual há duas fendas. Você dirá: o elétron pode

1. bater no anteparo e voltar; ou
2. passar pela primeira fenda do anteparo e ser detectado em uma região da tela bem em frente à primeira fenda, onde será registrado um “click”, ou ponto luminoso; ou
3. passar pela segunda fenda do anteparo e ser detectado em uma

região da tela bem em frente à segunda fenda, onde será registrado um “click”, ou ponto luminoso.

Note que, se assim for, após vários “clicks”, observaremos uma distribuição de pontinhos luminosos concentrada em regiões bem em frente às fendas, e praticamente nenhum ponto em outras regiões da tela detectora.

Não é isso que acontece, porém. Realizando o experimento, após um pequeno número de detecções, os “clicks”, ou pontinhos, no detector, parecem ser aleatórios. Mas, na medida em que o tempo passa, e mais e mais elétrons são lançados e detectados, observamos, na tela de detecção, *os pontinhos formarem um padrão de interferência exatamente como o da figura 9* – isto é, regiões claras onde chegam muitos elétrons alternados com regiões escuras, onde não chegam elétrons. Como se os elétrons fossem ondas.

Observe que isso acontece não obstante elétrons serem lançados e detectados – como pontinhos – *individualmente*. Repetimos, para que não haja dúvidas: *a imagem resultante da detecção de vários pontos luminosos individuais forma um padrão de interferência com algumas regiões mais claras, outras mais escuras*. Interferências construtivas e destrutivas, portanto.

O estudante deve ter claro para si o que isso significa: *cada elétron passa por ambas as fendas ao mesmo tempo*. Como se isso já não fosse suficientemente desconcertante, se fecharmos uma das fendas, e deixarmos apenas uma aberta, *o padrão ondulatório desaparece*, e os elétrons se distribuem na tela detectora como corpúsculos: concentrando-se, isso é dizer, na região do detector bem em frente à fenda. E se em seguida, abrirmos a fenda previamente fechada, novamente deixando ambas abertas, *o padrão de interferência volta a aparecer*.

“Contra-intuitivo” é uma expressão que não faz jus a quão extraordinariamente estranho é esse resultado. Por que o elétron – ou qualquer outra partícula material; o mesmo fenômeno é observado se o experimento for realizado com moléculas orgânicas de cerca de 1000 átomos – se importa se ambas as, ou apenas uma das, fendas está aberta?

A estranheza não para aí. Imagine que estejamos determinados em saber como o elétron se comporta quando passa pelas fendas; instalamos, então, um detector ao lado de uma das fendas, com o propósito de iluminá-la: assim, saberemos se o elétron passa por uma fenda, por outra, ou por ambas. E o que

acontece é o seguinte: se realizamos o experimento desse modo, *o padrão de interferência desaparece*. Como no caso em que apenas uma das fendas está aberta, os pontinhos luminosos que indicam a detecção de elétrons na tela se dispõem nas regiões bem em frente às fendas: como esperaríamos que fosse caso os elétrons fossem partículas.

O elétron, portanto, “parece se importar”, além de tudo, *em ser observado*. Nesse experimento, quando ambas as fendas estão abertas e nenhum detector disposto junto às fendas, o elétron parece passar por ambas as fendas, interferindo construtiva e destrutivamente *consigo mesmo* – “materializando-se como um pontinho” ao ser detectado; quando observado antes de passar pelas fendas, então, já aí “passa a se comportar” como partícula. Essa “dependência da medição” do comportamento corpuscular (discreto, não-linear) *versus* ondulatório (contínuo, linear) de objetos quânticos é parte do chamado “problema da medição”.

Como veremos, esse problema está no coração da teoria quântica. O estudante deve prosseguir a leitura atento ao fato de a teoria quântica buscar explicar, ou descrever, a realidade como ela se mostra, por exemplo, nesses experimentos.

2.2. Formalismo e paradoxos

Não obstante a teoria da relatividade exigir, como discutido acima, drástica transformação da visão de mundo clássica, a teoria, que unifica as noções outrora essencialmente distintas de espaço e tempo, não tardou em ser experimentalmente corroborada e amplamente aceita pela comunidade científica. O caminho que levou à teoria quântica foi, porém, completamente diferente. Ao passo em que a teoria da relatividade foi motivada por inconsistências teóricas entre o eletromagnetismo e o princípio de Galilei – note estar implícito, no teorizar de Einstein, o pressuposto metafísico segundo o qual a Natureza é inteligível e, por conseguinte, auto-consistente –, a teoria quântica foi motivada, como dissemos, pela intolerável inconsistência entre as predições da teoria clássica e os fatos observados. Até 1924, a chamada “velha teoria quântica” consistia em uma mistura confusa de conjecturas, regras *ad hoc* e problemas não-resolvidos. E de repente – entre o fim de 1925 e o início de 1926 – os físicos tinham à disposição *duas* teorias: a mecânica matricial e a mecânica ondulatória.

O formalismo atual da mecânica quântica – o chamado “formalismo do espaço de Hilbert da mecânica quântica” – é entendido, pela maioria dos físicos, como uma “síntese” das mecânicas matricial e ondulatória. Vamos, agora, revisar alguns dos fatos estruturais relevantes dessas teorias, a fim de entender a razão de ser de alguns dos problemas existentes no formalismo vigente.

A mecânica matricial foi desenvolvida por Werner Heisenberg, Max Born e Pascual Jordan entre junho e outubro de 1925. Heisenberg foi seu idealizador. Notando que o modelo do átomo de Bohr e a teoria clássica do movimento do elétron são incompatíveis, Heisenberg, sob influência do positivismo de Ernst Mach, procurou fazer uma teoria que descrevesse os resultados experimentais sem fazer referência à dinâmica do elétron: não observamos, pensou Heisenberg, a trajetória do elétron; não enxergamos – nem com nossos olhos, nem com a ajuda de aparatos experimentais – os orbitais atômicos. Concluiu Heisenberg: devemos construir uma teoria que nos permita calcular *diretamente* as quantidades físicas que são, em experimentos, *observadas* – como a energia, a intensidade e a frequência da radiação medida – sem fazer hipóteses sobre como é a Natureza no nível microscópico, inacessível aos nossos sentidos (veremos adiante como essa tentativa de “evitar a dinâmica” acabou por implicar em uma completa revolução na própria dinâmica).

O estudante é recordado de que, quando um elétron, preso em um átomo, passa de um orbital de energia maior (“mais externo”) para outro de energia menor (“mais interno”), o átomo emite energia de valor idêntico à diferença entre as energias dos orbitais envolvidos na transição. Para os inventores da mecânica matricial, esse fato sugeriu que a grandeza “energia” – hoje chamada “observável” – seja representada, na teoria quântica, por um objeto matemático que não é, como na física clássica, uma função. Os representantes matemáticos de grandezas quânticas devem ser, sob a ótica positivista de Heisenberg, Born e Jordan, objetos que dependem *de dois índices* – um para cada orbital. Você já deve ter encontrado um objeto com tais características no ensino médio: *matrizes*. Resumidamente, a mecânica matricial é um formalismo que reescreve as equações dinâmicas da física clássica, *substituindo funções por matrizes*. Esse procedimento levou à determinação de regras sistemáticas com as quais foi possível calcular, corretamente, as energias, frequências e intensidades da radiação emitida por átomos, quando elétrons passam de um orbital para outro.

Agora, a pergunta a ser feita é: que tipo de mundo a mecânica matricial descreve? As dificuldades interpretativas da mecânica quântica começaram já aí, no nascimento da teoria. Enquanto, na física clássica, os fenômenos empíricos motivaram a construção de objetos matemáticos (funções lineares, contínuas etc), no caso da teoria quântica, recorreu-se a uma matemática sofisticada, já existente, difícil de se interpretar fisicamente.

A mais imediata dessas dificuldades interpretativas advém do fato de matrizes não necessariamente comutarem. Você deve se lembrar de suas aulas de matemática no ensino médio: em certos casos, se A e B são matrizes, $A*B \neq B*A$ (em outras palavras, $(AB - BA) \neq 0$). No caso da teoria quântica, as “matrizes” (hoje chamadas “operadores”, por razões técnicas) \hat{X} e \hat{P}_x – representantes dos “observáveis” “posição” e “momento” – não comutam: $(\hat{X}\hat{P}_x - \hat{P}_x\hat{X}) = i\hbar$. O que esse fato matemático significa fisicamente?

Passemos à mecânica ondulatória. Erwin Schrödinger, motivado epistemologicamente por considerações completamente diferentes das acima descritas, e tendo em vista uma metafísica clara e realista para o mundo atômico, apresentou, em quatro artigos publicados entre janeiro de 1926, os fundamentos do que hoje chamamos “mecânica ondulatória”. Essa teoria não poderia ser mais diferente do que aquela proposta por Heisenberg, Born e Jordan. Schrödinger fora motivado pelos resultados de Louis de Broglie, historiador francês, que demonstrou teoricamente, em sua tese de doutorado, partículas materiais possuem propriedades ondulatórias (o experimento da dupla fenda com elétrons é posterior à publicação das ideias de de Broglie). Partindo do ponto de vista de que o elétron, em última instância, é *uma onda eletromagneticamente carregada* – seu aspecto corpuscular, manifesto em certos contextos experimentais, advém, de acordo com Schrödinger, do simples fato de nessas situações, a onda estar bem localizada no espaço, como em um “pacote” –, Schrödinger buscou determinar a equação correta – que deveria, note, ser *uma equação de onda* – que descrevesse o elétron como uma onda estacionária presa no átomo por um potencial dado pelo potencial coulombiano. A resposta,

$$i\hbar \frac{\partial \psi}{\partial t} + \frac{\hbar^2}{2m} \nabla^2 \psi - V\psi = 0, \quad (20)$$

onde i denota a unidade imaginária, \hbar a constante de Planck reduzida, m a

massa do elétron, e V o potencial ao qual o elétron está submetido, é a famosa “equação de Schrödinger”. Em posse dessa equação, Schrödinger calculou as possíveis energias de um elétron preso no átomo de hidrogênio. Os valores que ele obteve coincidiram com aqueles calculados a partir das regras *ad hoc* da “velha teoria quântica” e com os resultados experimentais.

Filosoficamente, vale notar que Schrödinger – como Einstein – partiu, em seu teorizar, do pressuposto metafísico segundo o qual a Natureza é inteligível, e do pressuposto epistemológico segundo o qual o objetivo da ciência consiste não apenas em descrever a Natureza, mas em explicá-la. O positivismo e o instrumentalismo de Heisenberg, Born e Jordan – e de muitos físicos que, àquela época, subscreviam ao positivismo de Mach – eram inaceitáveis para Schrödinger (e para Einstein, e para de Broglie).

O estudante deve notar que a metafísica que se segue da mecânica matricial e da mecânica ondulatória são consequência imediata dos pressupostos metafísicos à luz dos quais as teorias foram feitas. Podemos generalizar este fato nos termos do slogan: “metafísica para dentro, metafísica para fora”.

Schrödinger mostrou aos físicos não ser necessário apelar a uma “matemática diferente” para lidar com o mundo quântico (note que a equação de Schrödinger, como as equações da mecânica clássica, é uma equação diferencial⁹⁸). As “regras de quantização” da velha teoria quântica e da mecânica de matrizes – que se justificavam, reiteramos, apenas por serem capazes de *descrever*, mas não de explicar, os resultados experimentais – fizeram-se epistemologicamente inteligíveis à luz da mecânica ondulatória: a equação (20) só admite solução para valores discretos de energia.

A pergunta se apresenta de imediato: como decidir entre propostas tão diferentes para lidar com as mesmas questões? Levantou-se, em vista desse dilema, a hipótese: talvez não seja necessário “escolher” uma teoria e “rejeitar” a outra. Não seriam as regras da mecânica matricial apenas um método

⁹⁸Uma equação diferencial é uma equação que aparece na função na forma de derivadas. Note que a equação de Schrödinger depende de derivadas de x e t (posição e tempo respectivamente).

matemático diferente, mas fisicamente equivalente, de lidar com elétrons e outras partículas como ondas genuínas?

Assim propôs Schrödinger. Seu projeto, porém, não foi recebido com entusiasmo por seus oponentes, em particular por Heisenberg – e Schrödinger, trabalhando sozinho, não conseguiu levá-lo adiante. A visão de Schrödinger era extraordinária; a mecânica ondulatória, fascinante – mas não obstante o sucesso no caso do átomo de hidrogênio, havia muitos problemas que a teoria, naquele momento histórico de 1926, era incapaz de resolver. A visão do elétron como uma onda espalhada no espaço, por exemplo – traduzida matematicamente pelo ψ em (20) –, só funciona no caso particular em que o sistema é composto por apenas uma partícula. Quando consideramos um sistema com N partículas, a função de onda ψ vive em um espaço matemático abstrato de $3N$ dimensões. Ora: ondas reais vivem em um mundo com apenas três dimensões de espaço; logo, não é possível afirmar que ψ representa algo físico. Ademais, a teoria de Schrödinger também não consegue explicar como, não obstante o comportamento ondulatório do elétron manifesto em certos experimentos como o da dupla fenda acima discutido, em certas circunstâncias, o elétron se comporta manifestadamente como uma partícula. Se não são “pontos”, elétrons devem ter, nessas circunstâncias, extensão muito pequena. Mencionamos acima que Schrödinger pensava tratar esse caso com a ideia de “pacote de onda” (imagina a onda como um “envelope” muito bem localizado no espaço); essa noção não é estranha à ondulatória clássica. Não obstante, porém, a plausibilidade dessa hipótese, Schrödinger não conseguiu demonstrar, com exceção do caso particular, e muito simples, do oscilador harmônico, que uma onda que está concentrada em um pequeno volume do espaço em algum instante t não acaba se espalhando, “desfazendo o pacote”, instantes depois – o que impossibilita tratar a trajetória de um elétron livre como uma onda bem localizada no espaço e estável no tempo.

O primeiro passo na “resolução” dessas dificuldades foi dado por Max Born, um dos criadores, você deve se lembrar, da mecânica matricial. Recordamos o estudante: toda onda tem uma amplitude, a “altura” da oscilação temporal da onda. Born sugeriu – e essa ideia, central à teoria quântica, é hoje chamada “regra de Born” – que interpretássemos ψ não como uma onda física, como sugerira Schrödinger, o inventor da função de onda, mas *o quadrado de sua amplitude* $-\lvert\psi\rvert^2-$ como representando *a probabilidade de um observável*

ser medido (note aí a linguagem da mecânica matricial), *em um determinado experimento, com um determinado valor* (por exemplo, a probabilidade de medir o observável “energia” com valor 3,2 MeV para uma partícula em uma dada situação experimental é de 47%).

A função ψ é chamada, à luz do insight de Born, “onda de probabilidade”. O formalismo do espaço de Hilbert da mecânica quântica, que devemos a John von Neumann – identificado, hoje, com a própria mecânica quântica – faz uso da equação de Schrödinger, da regra de Born, e da linguagem centrada em “medições” e “observáveis” originárias da mecânica matricial. A combinação é, a um só tempo, matematicamente impecável, experimentalmente excelente, e filosoficamente incongruente.

Vejamos. Esquivando-nos de fazer uso de linguagem técnica, uma vez que supomos o estudante não estar familiarizado com a linguagem própria da física, apresentamos, de maneira extremamente simplificada, o procedimento de estudos de sistemas quânticos à luz do formalismo vigente. Para um sistema quântico qualquer, caracterizado por certo potencial V , determinamos, primeiro, as propriedades gerais do sistema, *resolvendo a equação de Schrödinger*. Obtemos, desse modo, os chamados “estados possíveis” do sistema (repetimos: os estados possíveis são soluções da equação). Note que há grandezas físicas que caracterizam o sistema, algumas das quais podemos medir experimentalmente (posição, energia, momento linear, momento angular, spin etc). Chamamos essas grandezas de “observáveis” do sistema. O estudante deve notar: o que determinamos experimentalmente é o valor numérico dos observáveis, detectados *quando o sistema é medido em um determinado estado*. Agora note o seguinte. Em linguagem matemática, dizemos que ψ é um *vetor*. Vetor é um objeto que mora em um espaço vetorial; você estudou esse objeto no ensino médio. Os vetores que você estudou, porém, provavelmente moram em um espaço vetorial identificado com o espaço euclidiano; ψ mora em um espaço vetorial mais complicado, o chamado “espaço de Hilbert”. Não entraremos em detalhes, o ponto para o qual chamamos a atenção é o seguinte: você deve se lembrar, de seus estudos no ensino médio, que se x_1 é um vetor, x_2 é um vetor, então, combinações lineares disso – por exemplo, $3x_1 + 1x_2$ – também é vetor. O mesmo vale para o objeto ψ . Se ψ_1 é vetor solução da equação de Schrödinger, se ψ_2 é vetor solução da equação de Schrödinger, então, o vetor $\Psi = a\psi_1 + b\psi_2$ também é vetor solução da equação de Schrödinger (impomos $a^2 + b^2 = 1$). Em

linguagem de física quântica, chamamos esse estado particular, “superposição”: uma superposição é um estado “que é soma de estados”. Observe o quão difícil é interpretar – e mergulhamos, agora, em águas filosóficas profundas – *o que é um estado de superposição: nós jamais medimos* isso, nós jamais observamos, *na realidade*, algo que podemos chamar “estado de superposição”. Com efeito, o que observamos, invariavelmente, em experimentos, é *um dos termos* da superposição: imagine que queiramos medir o spin⁹⁹ de um elétron na direção z; o observável “spin na direção z” é representado, simbolicamente, por \hat{S}_z ; o elétron possui spin 1/2 e podemos medi-lo ou para cima (estado ψ_1 com valor de spin $+\frac{\hbar}{2}$) ou para baixo (estado ψ_2 com valor de spin $-\frac{\hbar}{2}$). Antes da medição, o estado do elétron é

$$\Psi = \frac{1}{\sqrt{2}}\psi_1 + \frac{1}{\sqrt{2}}\psi_2 \quad (21)$$

mas *ao ser medido*, o elétron “colapsa” para ψ_1 ou para ψ_2 . Essa “ruptura” da dinâmica da função de estado que descreve o sistema, ruptura “causada” “pela medição”, consiste no famoso *problema da medição*: assim como no experimento da dupla fenda, o elétron parece passar por ambas as fendas ao mesmo tempo se não for observado, comportando-se, então, como uma onda, mas, se o observarmos, ele “colapsa” e se comporta como uma partícula, passando ou por uma fenda, ou por outra, teoricamente falando a função de onda descreve bem o comportamento do elétron antes de ser observado ou medido (chamamos a atenção do estudante para o fato de a equação de Schrödinger ser determinística, como as equações da física clássica), mas, *quando fazemos uma medição, a dinâmica determinista e linear de Ψ é subitamente interrompida*.

Há, portanto, dois “momentos” na descrição dinâmica de sistemas

⁹⁹O spin é uma propriedade de partículas elementares que diz respeito, basicamente, a como a partícula responde a um campo magnético. Considere a situação em que um elétron com spin “up” se desloca para cima devido à presença de um campo magnético: um elétron com spin “down” se deslocaria, nessa situação, para baixo, da mesma maneira com que uma partícula com carga negativa sofre uma força em sentido oposto àquela sentida por uma partícula com carga positiva na presença de um campo elétrico.

quânticos: antes da medição; depois da medição. A evolução do sistema *no tempo, antes de medirmos o sistema, é dada pela equação de Schrödinger. Na medição, aplica-se o chamado “postulado da projeção”*: a medição “*escolhe*” um dos estados possíveis (um dos termos da superposição), e o sistema “*colapsa*” para *aquele estado*.

Muitos problemas filosóficos se apresentam aí ao mesmo tempo. Em primeiro lugar, vimos que, de acordo com a regra de Born, não devemos interpretar Ψ como representando algo físico: o módulo ao quadrado da amplitude da função de onda representa, diz-nos o formalismo, a *probabilidade* de medirmos o sistema naquele estado, não o “estado verdadeiro” do sistema. Olhe novamente para a equação (21): as amplitudes de ψ_1 para ψ_2 são tais que a probabilidade de medir o sistema em um estado, ou em outro, é de 50% cada. Essa mera identificação da função de onda com “onda de probabilidade”, porém, é insatisfatória. Imagine que a função de onda (21) representa a situação experimental do elétron no experimento da dupla fenda. Se o elétron passa por uma fenda, ou por outra, os eventos “*elétron passa por fenda 1*” e “*elétron passa por fenda 2*” são independentes, certo? Você deve ter aprendido na escola que quando um evento – neste caso, o resultado de uma medição – tem mais de um resultado possível, a probabilidade de você obter um resultado *ou outro* é a soma aritmética da probabilidade de cada resultado: em (21), temos Prob $\frac{1}{2}$ para evento 1, Prob $\frac{1}{2}$ para evento 2, $\frac{1}{2} + \frac{1}{2} = 1$. Agora lembre-se do seguinte: o elétron só passa ou por uma fenda, ou por outra, *se o observarmos antes de ele passar pelo anteparo*: do contrário, o elétron passa por *ambas as fendas ao mesmo tempo*, e como vimos, *ele interfere consigo mesmo*: a probabilidade de o elétron ser observado em determinado ponto da tela detectora, *dada essa interferência, pode ser menor* do que seria no caso de o elétron ser observado antes de passar pelo anteparo (e ter “colapsado”). Na linguagem da teoria ondulatória, isso se passa porque ψ_1 interferiu destrutivamente com ψ_2 , o que explicaria, nessa linguagem, a “redução da probabilidade”. Mas ora! Se a função de onda é apenas uma ferramenta para calcular probabilidades, como pode ocorrer essa interferência? Probabilidades não “interferem” umas com as outras! Mais problemas surgem com essa constatação. Se a função de onda “tem algo de real”, como o fenômeno da interferência parece sugerir – portanto, ela não consiste, apenas, em uma medida de probabilidade –, então precisamos determinar o que é esse “elemento de realidade” que a função de onda capta.

Repetimos: a estrutura vetorial do espaço de Hilbert é tal que se ψ_1 e ψ_2 são estados, então, combinações lineares de ψ_1 e ψ_2 – os chamados “estados de superposição” – também são estados. No início desta seção sobre teoria quântica, fizemos a afirmação: *o mundo é quântico*. Note o que isso significa: a função de onda é capaz de descrever – ou deve ser, se a teoria está correta e é completa – *todo e qualquer sistema dinâmico na natureza*. Isso inclui, é claro, sistemas macroscópicos. Mas, repetimos, *não temos experiência de estados de superposição*.

Mostrar que o formalismo da teoria quântica, assim entendida, leva a absurdos é o ponto de Schrödinger quando ele discute, no artigo “Die gegenwärtige Situation in der Quantenmechanik” (“A situação atual da mecânica quântica”), publicado originalmente em 1935, o famoso experimento mental conhecido como “gato de Schrödinger”:

Pode-se até configurar casos muito ridículos. Um gato é trancado dentro de uma câmara de aço junto ao seguinte dispositivo (que deve ser protegido contra a interferência direta do gato): em um contador Geiger, há uma porção pequena de substância radioativa, *tão pequena*, que talvez, após uma hora, um dos seus átomos decaia, mas com igual probabilidade, talvez nenhum átomo decaia; se isso acontecer, o tubo contador libera uma descarga e através de um relé solta um martelo que, por sua vez, estilhaça um pequeno frasco com ácido cianídrico. Se deixarmos todo este sistema isolado durante uma hora, então diremos que o gato ainda vive *caso nenhum átomo tenha decaído*. A função ψ do sistema como um todo expressa essa situação contendo em si o gato vivo e o gato morto (perdoem a expressão) simultaneamente, ou dispostos em partes iguais [...]. (p. 812)

Encaremos a realidade. Não observamos superposições: observamos o gato *ou vivo ou morto* – não uma superposição disso. É preciso, por conseguinte: ou, (1) “esconder” as superposições, por assim dizer, em partes da realidade

onde elas não podem ser observadas, ou, (2) explicar *por qual razão* não temos experiência de superposições de maneira consistente com o formalismo da mecânica quântica, ou, (3) rejeitar a teoria, sob a afirmação de que ela faz previsões que não se verificam na realidade.

Ou estaria a teoria quântica correta, e nossa ideia de “realidade”, isto é, a maneira com a qual temos experiência de objetos macroscópicos – gato ou vivo, ou morto – errada?

Com essas considerações, você pode entender melhor o quão complicado, o quão bizarro, é o mencionado “problema da medição”. Você pode protestar: quando *medimos o sistema*, ele *colapsa* para um dos estados possíveis – por que isso não resolve a dificuldade?

Por um lado, o *mecanismo causal* que “provoca o colapso” é misterioso. Parece haver “algo sobre a medição” que escapa ao nosso entendimento epistemológico. Você pode achar ser capaz de contornar as dificuldades metafísicas – ou evitar a metafísica por completo, como quisera Heisenberg – ao olhar para a teoria com esse viés instrumentalista; “quando eu abro a caixa e olho para o gato, ele estará ou vivo ou morto, fim da história”. Considere a seguinte afirmação de Pascual Jordan, um dos inventores da mecânica matricial¹⁰⁰:

Observações não apenas perturbam o sistema, observações *produzem* a observação... Nós forçamos o elétron a assumir uma determinada posição... Nós mesmos *produzimos o resultado do experimento*.

Muito bem: mas e se você não olhar? Ou a realidade – gato vivo, gato morto, elétron em x , elétron em x' – só existe, ou passa a existir quando você observa? Atento a essa pergunta, considere, agora, o seguinte. Mencionamos, antes, que os operadores “posição na coordenada x ” e “momento na direção da coordenada x ” não comutam:

$$(\hat{X}\hat{P}_x - \hat{P}_x\hat{X}) = i\hbar;$$

¹⁰⁰Citado por Jammer 1974, p. 151.

O mesmo vale para as coordenadas y e z (posição e momento em coordenadas distintas comutam). Em continuação, fazemos notar que essas relações entre os operadores posição e momento – as chamadas “relações de comutação canônicas” – são fundamentais à teoria quântica; a pergunta, “qual seu significado?” é imediata. Responde Heisenberg: as relações de comutação nos levam ao chamado *princípio de incerteza*,

$$\Delta x \Delta p \geq \frac{\hbar}{2}, \quad (22)$$

onde Δx e Δp representam o “desvio padrão” da posição e do momento da partícula, respectivamente. Muito bem; como interpretar esse princípio? Continua Heisenberg: (22) significa existirem *limites* na nossa capacidade de *obter informação* de um sistema. Ou seja: se em um experimento, medirmos o observável \hat{X} e nosso resultado experimental for tal que o desvio padrão do experimento é Δx , então, o que podemos saber sobre o sistema, caso façamos uma medição de \hat{P}_x , não excede $\frac{\hbar}{2\Delta x}$. O estudante pode compreender o que está em jogo aqui, epistemologicamente, considerando o seguinte caso extremo: se medirmos a posição elétron com precisão, $\Delta x \approx 0$, e à luz de (22), *nosso conhecimento do momento do elétron é infinitamente pequeno* (incerteza infinitamente grande). De acordo com Heisenberg, essa interpretação de (22) é motivada *fisicamente*: quando medimos o elétron, *o perturbamos*. Para localizar o elétron, precisamos iluminá-lo com luz de comprimento de onda muito pequeno (alta frequência); quando isso acontece, a luz “bate” no elétron, alterando seu momento. Se, em contrapartida, usarmos luz de grande comprimento de onda, de modo a não “bater” no elétron e alterar, com isso, seu momento, não conseguimos, daí, localizá-lo no espaço. (22) codifica, diz Heisenberg, essas limitações epistêmicas consequentes da perturbação causada pela interação da medição; para medir o sistema, precisamos interagir com ele: logo, alteramos seu estado.

Você deve estar perplexo. Não foi dito, antes, na tentativa de evitar dificuldades ligadas à interpretação de estados de superposição, que é a medição que “cria” o estado? Falar em “perturbar o estado do sistema com a medição” certamente pressupõe a existência prévia do estado perturbado!

Advertimos o estudante: o problema de entender o papel da medição na mecânica quântica e o problema de entender a natureza do estado representado pela função de onda são apenas uma parcela muito pequena dos muitos mistérios que o mundo quântico nos apresenta. As chamadas “interpretações da mecânica quântica”, das quais você talvez já tenha ouvido falar, são uma tentativa de resolver os dilemas dessa teoria; as interpretações mais famosas são as chamadas “de Copenhague”, devidas a N. Bohr e W. Heisenberg, a dita “estocástica”, a “interpretação dos muitos mundos”, de H. Everett e J. Wheeler e a chamada “interpretação da onda-piloto”, ou “das variáveis ocultas”, de L. de Broglie e D. Bohm. Não trataremos, aqui, desse assunto; chegamos ao limite do que poderíamos apresentar em um número tão restrito de páginas, em nível introdutório, e sem fazer uso da linguagem formal da teoria. Fazemos notar, porém, que as interpretações não levam a consequências empíricas, e que não há consenso, nem entre físicos, nem entre filósofos, sobre qual interpretação é “a mais adequada”.

Não obstante as lacunas em nossa apresentação, inevitáveis em ensaios curtos e introdutórios, esperamos ter deixado claro ao estudante o quão hipotético é nosso conhecimento. A clareza esplêndida da mecânica clássica – forças agindo em corpos segundo as leis do movimento no espaço e no tempo absolutos – não mais existe. Procuramos indicar acima que a mecânica matricial e a mecânica ondulatória foram tentativas filosoficamente distintas de construir uma nova visão de mundo, à luz dos resultados experimentais, que demonstraram a falsidade da física clássica; essas visões foram “unidas” por von Neumann, de maneira matematicamente excelente, e empiricamente eficaz, no formalismo atual, o “formalismo do espaço de Hilbert da mecânica quântica”. Os experimentos que levaram à formulação da teoria ainda nos parecem misteriosos, porém. O mundo quântico, a natureza dual da luz e da matéria, continuam a desafiar a nossa compreensão. A tarefa filosófica de construir uma imagem de mundo que substitua a da física clássica ainda está em andamento.

Questões de revisão

1. Que problema Newton procurava resolver, ao postular o espaço e tempo absolutos?
2. O que define um referencial inercial? E um referencial não-inercial?

3. Qual o propósito do experimento do balde?
4. Qual a diferença entre movimento relativo e movimento absoluto, segundo Newton?
5. Por que, segundo Newton, os pontos do espaço (identificado matematicamente com \mathbb{R}^3) e os instantes do tempo (identificado matematicamente com \mathbb{R}) *existem*?
6. Como definimos “evento” em física?
7. Por que o espaço absoluto implica, de acordo com Leibniz, na existência de um número infinito de mundos possíveis empiricamente indistinguíveis?
8. Quais os postulados da relatividade restrita?
9. Qual a diferença epistemológica entre a noção de simultaneidade de observadores galileanos e a de observadores lorentzianos?
10. O que é o “paradoxo” no chamado “paradoxo dos gêmeos”?
11. O que é a regra de Born?
12. O que é um estado quântico?
13. Por que a interpretação da função de onda em termos de probabilidades é insatisfatória?
14. O que é um estado de superposição? É possível interpretá-lo fisicamente? Por quê?
15. O que é o princípio da incerteza de Heisenberg? Qual a incompatibilidade da interpretação desse princípio nos termos de “perturbação do sistema pela medição” com a ideia de que “é o observador que ‘produz’ o resultado do experimento”?
16. Por que representantes matemáticos de grandezas físicas na mecânica quântica são chamados “observáveis”?

Questões de discussão

1. O que significa dizer que as leis da física são invariantes por transformações de Galilei? Quais as consequências metafísicas desse fato para as noções de espaço e de tempo newtonianas?
2. Por que a realidade do movimento, de acordo com Newton, implica na realidade do espaço e do tempo?
3. Quais as motivações metafísicas das objeções de Leibniz à física de Newton?

4. Qual a incompatibilidade metafísica que Einstein identificou existir, na física pré-relativística, entre as leis da eletrodinâmica e as da física clássica? Em que sentido podemos dizer que o postulado da constância da velocidade da luz é parte do conteúdo do princípio da relatividade?
5. O que determina a possibilidade de influência causal entre dois eventos pontuais quaisquer A e B de duração instantânea?
6. Determine a velocidade de uma partícula que leva dois anos a mais do que leva a luz para percorrer 6 anos-luz de distância.
7. Uma barra mantém-se paralela ao eixo x de um referencial inercial S , movendo-se ao longo desse eixo com velocidade $0,630c$. O comprimento de repouso da barra é $1,70$ m. Qual é o tamanho da barra do ponto de vista de S ?
8. (a) Pode uma pessoa, em princípio, viajar da Terra até o centro galáctico (que está a cerca de 23.000 anos-luz de distância) num intervalo igual ao de uma vida normal? Explique usando argumentos da dilatação do tempo ou da contração dos comprimentos. (b) Que velocidade constante seria necessária para fazer essa viagem em 30 anos?
9. Um experimentador arma um mecanismo para disparar simultaneamente dois holofotes, um de flash azul, localizado na origem do seu referencial, e um de flash vermelho, em $x = 30$ km. Um segundo observador, movendo-se com a velocidade de $0,250c$ no sentido dos x crescentes, também vê os flashes. (a) Qual é o intervalo de tempo que ele registra entre os flashes? (b) Qual é o flash que, para ele, ocorre primeiro?
10. Quais as implicações metafísicas do fato de a transformada temporal t depender da coordenada temporal x e da velocidade v ?
11. Como o conceito de “espaço-e-tempo” difere, metafisicamente, do conceito de espaço-tempo?
12. Por que, de acordo com Putnam, os fatos da relatividade restrita implicam na refutação da tese metafísica segundo a qual o futuro é aberto?
13. Qual a diferença entre “passado, presente e futuro” de acordo com o eternalismo?
14. Descreva os fatos experimentais que levam à noção de “dualidade onda-partícula” em mecânica quântica. Como pode a energia de um fóton ser dada por $E = hf$, quando a simples presença da frequência f indica que a luz é uma onda?

15. Por que o experimento da dupla fenda com elétrons é impossível de ser explicado nos termos da física e ontologia clássicas?
16. Selecione a alternativa que apresenta as palavras que completam corretamente as lacunas, pela ordem, no seguinte texto relacionado com o efeito fotoelétrico:

“O efeito fotoelétrico, isto é, a emissão de _____ por metais sob a ação da luz, é um experimento dentro de um contexto físico extremamente rico, incluindo a oportunidade de pensar sobre o funcionamento do equipamento que leva à evidência experimental relacionada com a emissão e a energia dessas partículas, bem como a oportunidade de entender a inadequacidade da visão clássica do fenômeno. Em 1905, ao analisar esse efeito, Einstein fez a suposição revolucionária de que a luz, até então considerada como um fenômeno ondulatório, poderia também ser concebida como constituída por conteúdos energéticos que obedecem a uma distribuição _____, os quanta de luz, mais tarde denominados _____.”

- a) fótons – contínua – fótons
 - b) fótons – contínua – elétrons
 - c) elétrons – contínua – fótons
 - d) elétrons – discreta – elétrons
17. Uma placa de metal isolada emite fotoelétrons ao ser iluminada por luz ultravioleta, mas, depois de certo tempo, deixa de emití-los. Explique por quê.
 18. Que características epistemológicas e metafísicas distinguem a mecânica matricial da mecânica ondulatória?
 19. Por que não é possível interpretar a função de onda Ψ fisicamente?
 20. O que significa dizer que as grandezas físicas posição e momento de uma partícula comutam na física clássica, mas não na física quântica? Qual a razão matemática por trás disso? Qual a interpretação física desse fato? Qual a sua consequência epistemológica? Qual a sua consequência metafísica?
 21. Demonstre que se $\Psi_1(x, t)$ e $\Psi_2(x, t)$ são soluções da equação de Schrödinger dependente do tempo em uma dimensão, $\Psi_3(x, t) = C_1\Psi_1(x, t) + C_2\Psi_2(x, t)$ (onde C_1 e C_2 são constantes arbitrárias) também é solução.

22. Explique o que é o “colapso” da função de onda. Por que utilizamos esse termo na descrição de experimentos quânticos?
23. Qual o papel da medição na física quântica? Em que sentido esse aspecto da teoria é consequente da “síntese conceitual” da mecânica ondulatória e matricial?
24. Em que sentido o resultado do experimento da dupla fenda “depende do observador”? Qual a consequência metafísica disso?

Leituras recomendadas

Espaço, tempo, espaço-tempo e movimento:

Obras filosóficas que discutem esse tema em detalhe e profundidade são SKLAR, L. (1974) *Space, Time and Spacetime*. Berkeley: University of California Press, e DANTON, B. (2010) *Time and Space*. Montreal: McGill-Queen's University Press. Um livro elementar e divertido sobre a natureza do tempo é CALLENDER, C. & EDNEY, A. (2001) *Introducing Time: A Graphic Guide*. London: Icon Books. Leituras mais avançadas, filosoficamente muito ricas, mas que exigem conhecimento científico técnico do estudante são REICHENBACH, H. (1958) *The Philosophy of Space and Time*. New York, Dover; REICHENBACH, H. (1956) *The Direction of Time*. Berkeley: University of California Press; EARMAN, J. (1989) *World enough and space-time: absolute versus relational theories of space and time*. Cambridge: MIT Press. Em português, recomendamos fortemente a leitura de EINSTEIN, A. (1999) *A teoria da relatividade especial e geral*. Rio de Janeiro: Contraponto. Orig. em alemão: 1917, e de MARTINS, R. A. (2008) *Teoria da relatividade especial*. Campinas: Grupo de História e Teoria da Ciência.

Livros didáticos de física em língua portuguesa são FEYNMAN, R et al. (2008) *Lições de Física de Feynman* (3 volumes). Porto Alegre: Bookman, e NUSSENZVEIG, H. M. *Curso de Física Básica* (4 volumes). São Paulo: Blucher. Nas *Lições*, mecânica e relatividade são apresentadas no volume 1 e na obra do Prof. Nussenzveig, nos volumes 1 e 4, respectivamente.

Sobre fundamentos de mecânica quântica:

Em língua portuguesa, recomendamos fortemente PESSOA JR., O. (2006). *Conceitos de física quântica*. São Paulo: Livraria da Física. Esse livro discute conceitos e problemas filosóficos em mecânica quântica à luz de experimentos; a leitura exige conhecimento elementar prévio da teoria. Material de cunho mais introdutório foi escrito pelo Professor Pessoa Jr. para o site *Vya Estelar*; vide <http://www.fflch.usp.br/df/opessoa/Vya-Quantica-Tudo.pdf>. Traduções para o português de obras importantes, e de fácil leitura, sobre fundamentos de física quântica são HEISENBERG, W. (1996) *A parte e o todo*, trad. Vera Ribeiro, Contraponto, Rio de Janeiro. Orig. em alemão: 1969. Versão de Portugal: *Diálogos sobre física atômica* (1975); BOHM, D. (2008) *Totalidade e a ordem implicada*, trad. T. Lorent, Madras, São Paulo. Publicada em inglês em 1980, tem uma tradução anterior pela Cultrix, 1992; BOHR, N. (1995) *Física atômica e conhecimento humano*, trad. Vera Ribeiro, Contraponto, Rio de Janeiro.

Em língua inglesa, recomendamos JAMMER, M. (1974) *The philosophy of quantum mechanics*. Wiley: Nova Iorque. É um dos livros de cabeceira de todo filósofo da teoria quântica; é preciso domínio técnico da teoria para entender o texto. ALBERT, D. (1994) *Quantum mechanics and experience*. Harvard University Press. Altamente recomendado para estudante de graduação em filosofia, a apresentação da teoria quântica, aí, é tão elementar quanto possível. HEISENBERG, W. *The Physical Principles of the Quantum Theory*. Dover: New York. Um dos grandes clássicos, escrito por um dos grandes nomes da teoria quântica, recomendamos fortemente esse livro para estudantes com conhecimento elementar de física quântica. MAUDLIN, T. (2011) *Quantum Non-Locality and Relativity*. Cambridge: Cambridge University Press. De nível mais avançado, esse livro discute o teorema de Bell e implicações, do qual não tratamos neste ensaio. O estudante já deve ter conhecimento de teoria quântica e de filosofia da mecânica quântica para entendê-lo.

O volume 3 de *Lições de Física de Feynman* é recomendado como livro didático de introdução à teoria da mecânica quântica. Recomendamos, também, fortemente, GRIFFITHS, D. (2011) *Mecânica quântica*. Londres: Pearson Education.

Referências

- EINSTEIN, A. (1905a) "Ueber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt". *Annalen der Physik* 17: 132-148.
- _____. (1905b) "Zur Elektrodynamik bewegter Körper". *Annalen der Physik* 17: 891-921.
- GALILEI, G. (1632; 2011) *Diálogo sobre os Dois Máximos Sistemas de Mundo Ptolomaico e Copernicano*. Editora 34: São Paulo. Tradução de P. R. Mariconda.
- JAMMER, M. (1974) *The Philosophy of Quantum Mechanics*. Wiley: New York.
- KRAGH, H. (2010) *Quantum Generations*. Princeton University Press: Princeton.
- LANGEVIN, P. (1911) "L'évolution de l'espace et du temps". *Scientia* 10: 31-54.
- PUTNAM, H. (1967) "Time and Physical Geometry". *The Journal of Philosophy* 64(8): 240-247.
- SCHRÖDINGER, E. (1935) "Die gegenwärtige Situation in der Quantenmechanik". *Die Naturwissenschaften* 23: 807-812.

8

Filosofia da Biologia

Sérgio Farias de Souza Filho

1. Introdução: o que é a filosofia da biologia?

Ao contrário das ciências naturais e humanas, a filosofia não se define pelo seu objeto de investigação. Isto é, enquanto a física tem como objeto de investigação a realidade física e as ciências sociais têm como seu objeto a realidade social, a filosofia não tem como seu objeto uma parte ou aspecto da realidade em contraposição a outra. O que ocorre é que a filosofia investiga a *realidade como um todo*. Dito de outra forma, não há nada na realidade que não seja objeto de reflexão filosófica. Mas se tudo pode ser objeto de reflexão filosófica, então as próprias teorias e explicações científicas acerca de aspectos da realidade também podem ser objeto de investigação filosófica. É assim que surge não apenas a filosofia da ciência enquanto tal, mas também a filosofia de cada ciência particular – filosofia da matemática, da física, das ciências sociais etc.

Diante disto, parece ser fácil caracterizar a filosofia da biologia. Trata-se afinal da área da filosofia que tem como objeto de investigação uma ciência particular, a saber, a biologia (do grego, *bio*, vida + *logia*, estudo). Esta é a ciência natural que tem como objeto os seres vivos e processos vitais. Ela desenvolve teorias para explicar a natureza, origem e desenvolvimento de organismos (e.g., animais, plantas e bactérias) e processos vitais (e.g., reprodução, adaptação e metabolismo). Tal caracterização da filosofia da biologia serve como primeira aproximação, mas revela-se insuficiente. Quais são os problemas filosóficos peculiares à filosofia da biologia? Em que consiste a reflexão filosófica acerca das práticas e teorias das ciências biológicas?

Pode-se dizer que a filosofia da biologia enquanto disciplina filosófica desenvolvida e distinta surge na segunda metade do século XX, gerando enorme

literatura desde então. Atualmente ela é uma das áreas mais vibrantes e excitantes da filosofia contemporânea. Em sentido amplo, a filosofia da biologia comporta *três tipos* de investigações filosóficas (OBENDAUGH & GRIFFITHS, 2020). No *primeiro tipo*, se encontram aquelas investigações que tratam de problemas e teses típicas da filosofia da ciência, mas no contexto das ciências biológicas – realismo científico, causalidade, explicação científica, lei científica etc. No *segundo tipo*, estão aquelas investigações filosóficas que recorrem aos resultados e teorias biológicas para lidar com problemas filosóficos tradicionais das mais diversas áreas da filosofia, como metafísica, ética, filosofia da mente etc. Por fim, no *terceiro tipo*, estão aquelas investigações filosóficas que lidam com problemas teóricos, metodológicos e conceituais específicos da biologia tais como os conceitos biológicos de função, adaptação, espécie, unidade da seleção natural etc. Embora distintos, estes três tipos de investigações filosóficas estão interligados uma vez que os resultados de uma investigação de um certo tipo podem influenciar o desenvolvimento de investigações de outro tipo.¹⁰¹

Nesta introdução à filosofia da biologia, focarei no terceiro tipo de investigação filosófica. No que se segue, irei apresentar e discutir problemas teóricos e conceituais acerca das práticas e teorias científicas da biologia. Uma vez que a biologia é dividida em diversas áreas (biologia evolutiva, genética, do desenvolvimento etc.), segue-se que as práticas e teorias científicas provenientes de cada área podem ser sujeitas à investigação filosófica. Contudo, devido à limitação de espaço disponível e ao fato de esta ser uma introdução à filosofia da biologia que não pressupõe conhecimento prévio da área, tive que focar em alguns problemas da filosofia da biologia e deixar outros de lado. Darei especial atenção a uma subárea específica da biologia, a saber, a *biologia evolutiva*. A principal razão para tanto se assenta no fato de que a filosofia da biologia evolutiva é historicamente o cerne dos debates contemporâneos da filosofia da biologia, tendo experimentado o maior grau de desenvolvimento e amadurecimento dentre todas as suas áreas. Por fim, a propaganda corrente é

¹⁰¹Por exemplo, uma concepção de função biológica influencia como se lida com o tradicional problema da intencionalidade na filosofia da mente. De fato, as chamadas teorias teleológicas da representação mental (“teleosemântica”) procuram solucionar este problema, apelando para a assim chamada concepção etiológica de função biológica (SOUZA FILHO, 2013). Abordarei o tópico filosófico da função biológica mais adiante.

que “nada na biologia faz sentido exceto à luz da evolução”¹⁰² (DOBZHANSKY, 1973).

A biologia evolutiva é o campo da biologia que investiga os processos evolutivos, aqueles que produziram a diversidade de seres vivos na Terra. Assim, a filosofia da biologia lida com problemas teóricos e metodológicos acerca das práticas e teorias da biologia evolutiva. Mas que problemas filosóficos são estes e em que estes diferem dos problemas científicos da biologia evolutiva? A natureza da distinção entre problemas científicos e filosóficos é em si um problema metafilosófico que está para além do escopo da presente introdução lidar, mas a caracterização a seguir é suficiente como primeira aproximação.

Problemas filosóficos da biologia tratam dos aspectos mais fundamentais das ciências biológicas, cujas respostas os biólogos geralmente não refletem a respeito em suas práticas científicas. Por exemplo, biólogos frequentemente adotam um discurso teleológico, que recorre à noção de função biológica, a fim de descrever a natureza de um organismo e seus traços fenotípicos (i.e., qualquer atributo observável de um organismo como altura, forma ou cor da pele). Em sua prática científica, os biólogos assumem que traços e mecanismos biológicos têm funções biológicas, mas geralmente não refletem fundamentalmente acerca do que seria precisamente uma função biológica. Evidentemente, os biólogos podem posteriormente refletir filosoficamente acerca da natureza das funções biológicas, mas via de regra suas práticas científicas não exigem tal empreendimento. O mesmo ocorre com outras noções como adaptação, espécie etc. Por fim, a filosofia da biologia também investiga questões como se de fato há leis naturais na biologia, se estados biológicos são redutíveis a estados físicos etc. Evidentemente, biólogos que investigam os fundamentos teóricos da biologia evolutiva podem também se interessar por problemas como estes – de fato, muitos o fazem – e é justamente por isso que a fronteira entre os campos da filosofia da biologia e dos fundamentos da biologia se torna por vezes porosa. De todo modo, o que ocorre é que em suas práticas científicas os biólogos estão mais focados em fazer descobertas empíricas e desenvolver teorias e modelos do que analisar filosoficamente os fundamentos de sua ciência.

¹⁰²Todas as traduções de citações em língua inglesa são de minha autoria ou foram retiradas das traduções para o português de edições publicadas no Brasil.

O objetivo desta introdução é apresentar ao leitor o estado da arte da filosofia da biologia através de alguns dos principais problemas filosóficos que moldam o debate contemporâneo nesta área. Na próxima seção, faço uma apresentação da teoria da evolução pela seleção natural, tal como estabelecida na biologia evolutiva contemporânea e avalio alguns de seus aspectos filosoficamente relevantes. Na terceira seção, trato do problema da teleologia e da função biológica. Na quarta seção, trato do problema taxonômico e ontológico da classificação das espécies. A quinta seção conclui esta introdução. A sexta e última seção contém questões de revisão, questões de discussão e sugestões de leitura.

2. Evolução

A Terra é habitada por organismos com os mais variados traços e dos mais variados tipos: árvores, sapos, macacos, peixes, abelhas, bactérias, protozoários... e finalmente nós, os seres humanos. Estima-se que ao longo da história, a Terra já foi habitada por bilhões de espécies (LEWONTIN, 1978). Diante de tão espetacular variedade de espécies, um problema se impõe: qual a origem de todas essas espécies? A tradição judaico-cristã nos conta que Deus criou o céu e a terra. Numa interpretação literal da bíblia, isso significa que Deus criou separadamente todas as espécies, incluindo os seres humanos que por sua vez foram por ele criados à sua imagem e semelhança. O criacionismo estabelece então uma resposta simples e direta para o problema da origem das espécies: a riqueza e variedade das espécies se explica pelo ato de criação divina. Deus criou cada espécie e por isso há tantas vagando na Terra.

A *teoria da evolução pela seleção natural* pôs abaixo o criacionismo e outras crenças então enraizadas no senso comum do homem ocidental quanto à origem das espécies. Em meados do século XIX, Charles Darwin e Alfred Russel Wallace, de maneira independente, desenvolveram a teoria da evolução pela seleção natural, embora Darwin tenha proposto a sua versão mais desenvolvida em sua obra-prima *A Origem das espécies por meio de seleção natural* (DARWIN & WALLACE, 1858; DARWIN, 1859 [2018]). A teoria estabeleceu uma reviravolta em nossa compreensão do mundo natural e é comumente considerada como o maior marco na história da biologia.

Primeiramente, o que é a evolução? Nada mais que a *modificação*

gradual dos traços de uma população biológica ao longo do tempo. A evolução consiste na mudança gradual da população ao longo das gerações. A teoria da evolução estabelece que as espécies não surgiram prontas, mas são o resultado de um longo e tortuoso processo evolutivo. O mundo natural não é constante. Ora, isto por si só já contraria a tese criacionista segundo a qual Deus criou as espécies já prontas, ou seja, que as espécies seriam imutáveis após o ato de criação divina. Mas este é apenas o início desta divergência.

A teoria da evolução também estabelece a *ancestralidade comum* – as espécies atuais descendem de um ou mais ancestrais comuns. Em última instância, todas as espécies atuais descendem de um mesmo ancestral. Isto implica que organismos tão díspares como seres humanos, bactérias, abelhas, baleias e aves possuem um mesmo ancestral, se voltarmos o suficiente em suas respectivas árvores evolutivas. Na chamada “árvore da vida”, há um último ancestral comum do qual todos os seres vivos descendem. Ora, isto vai de encontro a tese criacionista que Deus criou separadamente cada espécie. Mas como ocorreu este processo evolutivo que se iniciou com este ancestral comum e desaguou em todas as espécies que atualmente habitam a terra?

A *seleção natural* é o mecanismo ou meio pelo qual a evolução ocorre. Em sua formulação mais aceita na biologia contemporânea, tal como desenvolvida por Richard Lewontin, a teoria da evolução pela seleção natural estabelece três princípios gerais (LEWONTIN, 1970):

Princípio da variação fenotípica: diferentes indivíduos de uma dada população possuem diferentes traços.

Princípio da aptidão diferenciada: diferentes fenótipos têm diferentes proporções de sobrevivência e reprodução em diferentes ambientes.

Princípio da hereditariedade: as variações fenotípicas são hereditárias – os descendentes em média se assemelham mais aos seus genitores que a outros indivíduos (a aptidão é herdada).

Estes três princípios são suficientes para garantir a ocorrência da evolução pela seleção natural. Sempre que satisfeitos, haverá processo evolutivo – a seleção natural é o resultado inevitável da interação destes três princípios (STERELNY & GRIFFITHS, 1999, pp. 31-2).

Primeiramente, note que se não houver variação de traços biológicos

numa população, não há como haver seleção natural. Um processo de seleção consiste no favorecimento de um traço em detrimento do outro – se não há variação, simplesmente não há o que ser selecionado. Em segundo lugar, algumas variações fenotípicas estarão mais aptas para lidar com os problemas apresentados pelo ambiente do que outras (alimentação, predadores, caça, acasalamento etc.). Conseqüentemente, os organismos portadores de tais variações têm maior probabilidade de sobreviver para reproduzir (ou ao menos reproduzir de maneira mais eficaz) que os organismos portadores das variações concorrentes. Ou seja, a variação fenotípica resulta em uma variação de *aptidão* (*fitness*) – a habilidade de um traço ou organismo para sobreviver e se reproduzir. Por fim, as variações mais aptas devem ser hereditárias para que aumente a proporção das futuras gerações com as mesmas. A hereditariedade é necessária posto que em sua ausência a proporção da população com aquelas variações mais aptas permaneceria a mesma; portanto não haveria modificação evolutiva gradual ao longo das gerações. Uma vez satisfeitos, estes três princípios gerais garantem a ocorrência da seleção natural. Esta seleciona os organismos mais aptos e a hereditariedade de seus traços garante a mudança evolutiva na população descendente. Estes três princípios predizem que os traços com maior aptidão irão se tornar mais comum na população.¹⁰³

A seleção natural é uma ideia simples, porém profundamente revolucionária. Para ilustrar seu poder e generalidade, considere uma população de girafas numa savana. Por que girafas têm pescoços longos? A resposta direta é que girafas desenvolveram pescoços longos para que pudessem alcançar as folhas nos topos das árvores que os outros animais não podiam alcançar e assim pudessem se alimentar. Já a resposta completa é mais complexa. Na população anterior a este processo evolutivo, havia uma variação no comprimento dos pescoços das girafas: uns mais curtos, outros mais longos (variação fenotípica). Ocorre que pescoços longos tornam as girafas mais aptas ao ambiente por facilitarem a alimentação. Girafas de pescoços longos, portanto, tendem a sobreviver e se reproduzir em maior proporção que girafas de pescoços curtos (aptidão diferenciada). Por fim, o comprimento do pescoço é um traço hereditário

¹⁰³Há um profundo debate entre biólogos e filósofos sobre quais são as condições necessárias e/ou suficientes para haver seleção natural. O que ocorre é que estes três princípios são os mais populares, cf. GILDENHUYS, 2019.

– os descendentes em média têm o tamanho do pescoço mais semelhante ao tamanho do pescoço de seus genitores em contraste com outros indivíduos da população (hereditariedade). O resultado da interação destes três fatores é que, ao longo das gerações, as girafas descendentes passam a ter um pescoço em média maior que o de seus ancestrais. A evolução do comprimento dos pescoços das girafas é explicada pela seleção daqueles pescoços com comprimentos mais aptos.¹⁰⁴

Contudo, o processo evolutivo não depende apenas da ocorrência destes três fatores – variação fenotípica, aptidão diferenciada e hereditariedade. Para haver mudança adaptativa, tal como o crescimento do pescoço das girafas, é preciso haver *seleção cumulativa*. A mudança seletiva não ocorre em um único episódio de seleção, de uma vez por todas, mas é o resultado de episódios de seleção em série. A evolução é um longo processo cumulativo que ocorre ao longo de muitas gerações através de uma sequência colossal de episódios de seleção. As mudanças evolutivas possuem um tal grau de complexidade que não podem ser o resultado de alguns poucos episódios de seleção. Segundo Richard Dawkins, a seleção cumulativa é a única explicação bem-sucedida para a complexidade do design dos organismos (DAWKINS, 1986 [2001]).

Por fim, é preciso salientar o *caráter universal* dos três princípios acima elencados para haver evolução pela seleção natural – variação fenotípica, aptidão diferenciada e hereditariedade. Note que eles não mencionam organismos. O que ocorre é que a teoria da evolução deve explicar a evolução não apenas de organismos, mas deve ser suficientemente geral para explicar também a evolução das espécies, dos genes (em breve tratarei dos mesmos), de outras moléculas dentro dos próprios organismos. A teoria da evolução deve ser expressa não apenas como se aplicando a este ou aquele organismo, nem mesmo a todos os seres vivos, mas como uma teoria universal que se aplica a todas as coisas que se reproduzem desde que tenham variação hereditária e aptidão diferenciada (ROSENBERG & McSHEA, 2008, pp. 19-20).

¹⁰⁴ Aquisaliento que recorri ao caso da girafa com fins meramente ilustrativos. Esta explicação da evolução do pescoço das girafas não exclui a possibilidade delas terem desenvolvido pescoços longos por outras razões. Por exemplo, para intimidar predadores ou mesmo outras girafas, em contextos competitivos (ROSENBERG & McSHEA, 2008, p. 17).

Contudo, dois problemas se impõem para explicar a evolução pela seleção natural. O primeiro é que esta depende fundamentalmente de uma forte hereditariedade entre os traços dos genitores e os traços dos descendentes (princípio da hereditariedade). Porém, como isso é possível? Qual o mecanismo pelo qual a hereditariedade ocorre? Este foi um dos principais problemas que ameaçaram a viabilidade da teoria da evolução tal como originalmente proposta por Darwin. O segundo problema está no fato que a seleção natural também exige uma *variação de traços* nas populações sujeitas ao processo evolutivo (princípio da variação fenotípica). Contudo o mecanismo da seleção natural age justamente em sentido contrário, continuamente homogeneizando a população: a variação é reduzida através do favorecimento dos traços mais aptos e da eliminação dos traços menos aptos. Isto implica que para haver processo evolutivo por um longo período, é necessário haver uma contínua oferta de variação fenotípica. Mas o que garante tamanha oferta? A resposta para ambos os problemas está no *gene*, a unidade da hereditariedade, cuja teoria era desconhecida por Darwin.

O desenvolvimento da teoria genética constituiu uma das maiores evidências científicas para a teoria da evolução. A transmissão hereditária ocorre através dos genes que são compostos de ácidos nucleicos e o fornecimento “infinito” de variação fenotípica é garantido por dois fatores, a *recombinação* e a *mutação* genética (DAWKINS, 1976 [2007]). A maioria dos organismos se desenvolvem a partir de uma única célula. No caso da reprodução sexuada, esta célula é o zigoto, formado pela união de dois gametas: um masculino (o espermatozoide) e outro feminino (o óvulo). Os gametas são constituídos por genes que são transmitidos para os descendentes e consigo transmitem os traços de seus genitores. É assim que os descendentes *herdam* os traços de seus genitores. Ocorre que os genitores transmitem apenas metade de seus genes para seu gameta. A escolha destes genes é dita *aleatória* na medida em que o processo de divisão não leva em conta qualquer necessidade adaptativa. Isto gera uma recombinação constante de genes e é o primeiro fator para haver oferta contínua de variação fenotípica: cada organismo é o resultado de uma nova combinação genética. O segundo fator é a mutação genética – os genes não são transmitidos com absoluta fidelidade, mas são passíveis de modificação. Por vezes há mutação quando o material genético é copiado, de modo que o zigoto resultante da união dos gametas masculinos e femininos conterá genes

que originalmente não pertenciam a nenhum dos dois gametas. Assim, a mutação genética também contribui para uma oferta incessante de variação de traços sobre a qual age a seleção natural.

A biologia evolutiva contemporânea surge a partir da chamada *síntese moderna*, desenvolvida por R.A. Fisher, J.B.S. Haldane, Sewall Wright e outros na década de 1930 (TABERY, 2019). Ela integrou a teoria da evolução com a biologia genética, estabelecendo a evolução como sucessivas modificações na frequência genética da população através da seleção natural. Desde então, a síntese moderna tem sido refinada e alguns de seus aspectos tem sido postos em questão tal como sua tese de que o único mecanismo responsável pela evolução é a seleção natural (argumenta-se que outros fatores podem ser importantes no processo evolutivo tal como o acaso e fatores não genéticos no processo hereditário).

Por fim, uma última observação antes de passar para a próxima seção. Muitos que atualmente se opõem à teoria da evolução, especialmente aqueles de inspiração criacionista, costumam sublinhar que a teoria da evolução seria “apenas uma teoria” entre tantas teorias que procuram explicar a origem das espécies. Com isto, querem dizer que a teoria da evolução não goza de qualquer status epistêmico especial, mas seria uma teoria com o mesmo poder explicativo que as demais. Nada poderia ser mais falso. A teoria da evolução é um fato estabelecido na comunidade científica e há uma quantidade colossal de evidências que demonstram sua comprovação científica (para uma compilação das mesmas, cf. COYNE, 2009 [2014]). Toda teoria científica pode evidentemente em última instância ser posta em dúvida e isto não seria diferente com a teoria da evolução. Ocorre que esta goza de tamanho grau de comprovação científica que a pôr em dúvida constitui um ceticismo dos mais radicais quanto ao sucesso científico.

3. Função

Teleologia (do grego *telos*, “fim” + *logia*, “estudo”) é a explicação de um fenômeno a partir de sua função ou propósito. Ciências naturais como a física ou a química são destituídas de explicações teleológicas, suas teorias explicam seus respectivos fenômenos naturaissem apelar para qualquer finalidade ou propósito. Estas ciências não assumem que os fenômenos naturais que

constituem os seus domínios de investigação tenham qualquer fim. A biologia, contudo, é uma estranha no ninho: uma ciência natural ubíqua de explicações teleológicas. Este contraste entre a biologia e as outras ciências naturais levanta diversos problemas filosóficos. Por que a biologia difere das demais ciências naturais? Qual a razão para que ela supostamente não possa se dar ao luxo de dispensar explicações teleológicas? Por fim, o que é esta *finalidade ou propósito biológico* tão postulado em explicações biológicas? Nesta seção, apresentarei o problema da teleologia na biologia.

Explicações teleológicas se caracterizam por explicar a presença, ocorrência ou natureza de um fenômeno pelo fim para o qual este contribui. Na física e química contemporâneas, explicações teleológicas foram descartadas como cientificamente válidas, por inverterem a relação de causa e efeito: a teleologia assume que um efeito pode determinar sua própria causa. Mas ora, isto *inverte* a relação causal— são os efeitos que se seguem às suas causas, não o contrário. Nas palavras de Spinoza, “essa doutrina finalista inverte totalmente a natureza, pois considera como efeito aquilo que é realmente causa e vice-versa [i.e., considera como causa aquilo que é realmente um efeito]. Além disso, converte em posterior o que é, por natureza, anterior.” (SPINOZA 2009 [1677], II/80). Na sinuca, o efeito da bola de bilhar caindo na caçapa se segue à causa do taco arremessando a bola em direção à caçapa; não é a bola encaçapada que determina o taco a arremessando. A explicação mecânica para o movimento da bola não é, portanto, teleológica, mas uma descrição causal da sequência de efeitos que levam a bola a ser encaçapada. Agora considere a seguinte explicação biológica: o coração bombeia sangue pelas veias do corpo, porque o seu *propósito biológico* é precisamente bombear sangue. A explicação para o coração bombear sangue é que esta é a sua finalidade. Note como esta explicação recorre a um *evento futuro* a fim de explicar um *evento presente*. Este padrão explicativo é o que caracteriza as explicações teleológicas.

O problema que se impõe, portanto, é por qual razão a biologia explica o coração bombeando sangue apelando para a finalidade do coração, enquanto a física não explica o movimento da bola de bilhar apelando para a sua finalidade. Por que, apenas no primeiro caso, a postulação de uma função é supostamente justificada, mas não no último? Generalizando, qual a razão para ser proibido que explicações físicas façam referências a eventos futuros, enquanto, no caso de explicações biológicas, isto é supostamente justificado? É preciso fornecer

uma explicação para como traços biológicos podem ter qualquer função ou propósito. Afinal, recorrer a uma noção teleológica para desenvolver explicações biológicas, sem fundamentar em que a mesmas consistem, põe em risco a viabilidade dessa própria explicação. Aqui, poder-se-ia recorrer novamente à ação divina: Deus dotou cada traço biológico com uma função e é em virtude disso que os traços possuem as funções que efetivamente possuem. O problema desta abordagem para a teleologia biológica é sua incompatibilidade com o método naturalista das ciências naturais, que proíbe explicações científicas de recorrerem às entidades sobrenaturais como divindades (JACOBS, 2020). Sendo a abordagem sobrenatural uma carta fora do baralho, o que então poderia fundamentar a teleologia biológica?

A propaganda corrente é que explicações teleológicas são indispensáveis na biologia, devido ao papel explicativo, advindo da *prática de atribuição funcional* por parte dos biólogos (NEANDER, 2018; WALSH, 2008). A atribuição de uma função a um traço de um organismo dá origem a uma explicação parcial da existência ou da presença deste traço, na medida em que identifica a contribuição biológica do traço para o organismo em questão. A atribuição da função de bombear sangue ao coração explica o porquê de primatas terem coração; a atribuição da função de voo às asas explica o porquê de aves terem asas etc. Este papel explicativo da atribuição de funções biológicas é supostamente indispensável; explicações teleológicas na biologia nada mais são que *explicações funcionais*. Mas o que é uma função biológica? Como elas podem supostamente fundamentar explicações teleológicas quanto à presença de traços nos organismos?

A teoria mais influente da função biológica é a *teoria etiológica* (WRIGHT, 1973; MILLIKAN, 1989; NEANDER, 1991), também denominada *teoria dos efeitos selecionados*. Ela pode ser resumida no seguinte mote: a função de um traço é seu *efeito selecionado*. Traços de um certo tipo passaram por um *processo de seleção* no qual um certo efeito foi selecionado; como resultado, o traço passou a ter este efeito como sua função biológica. O processo de seleção é o que determina a função biológica. Como não poderia deixar de ser, o processo de seleção paradigmático é o da evolução pela seleção natural: a função do coração é bombear sangue porque corações ancestrais foram selecionados *para* bombear sangue. Em suma, é a *história de seleção* dos ancestrais do traço que determina sua função biológica.

Aqui, um esclarecimento faz-se necessário. É comum, quando falamos na função de um objeto, recorrermos às intenções do *designer* por trás daquele objeto. Este é o caso de artefatos: o propósito da lâmpada é iluminar o ambiente, porque seu designer a projetou *para* iluminar o ambiente. A função de artefatos depende constitutivamente da intenção do designer, isto é, de sua mente. Ora, a função biológica não é constituída por qualquer estado mental. Para a teoria etiológica, a seleção natural é um processo que independe do olhar do biólogo ou do observador. Antes mesmo da humanidade surgir na face da Terra, as populações já evoluíam através do mecanismo da seleção natural. Em suma, a função biológica de qualquer traço não é determinada pelas intenções de qualquer designer (e.g., Deus), ao passo que a função do artefato é determinada pelas intenções do artífice. Funções biológicas e de artefatos se distinguem em virtude de suas *etiologias*, i.e., suas origens históricas.

A teoria etiológica estabelece que a função de um item Z é X, se itens do mesmo tipo que Z foram selecionados para fazer X. Esta fórmula geral é aplicável não apenas às funções biológicas, mas também às funções de artefatos. Aplicando-a especificamente ao caso da evolução pela seleção natural, temos a seguinte definição formal (NEANDER, 1991):

Função Etiológica: o traço X do organismo O tem a função de fazer aquilo que itens do tipo X fizeram que contribuiu para a aptidão dos ancestrais de O e que causou a seleção do genótipo cuja expressão fenotípica é X.

Asas têm a função de voo, porque asas ancestrais contribuíram para a aptidão de aves ancestrais, o que causou a seleção do genótipo cuja expressão fenotípica são as asas.

Diversas são as virtudes da abordagem etiológica da função biológica. A primeira virtude é que ela sucede em fundamentar como atribuições funcionais podem parcialmente explicar a presença dos traços nos organismos. Um traço ter a função de fazer algo depende constitutivamente de seus traços ancestrais terem sido selecionados para fazerem isto. O efeito selecionado de um traço explica por que organismos possuem este traço, na medida em que este existe no presente, em virtude de seus ancestrais terem sido selecionados pelo que fizeram no passado. O que explica a existência ou manutenção de um traço em uma dada população é o *mecanismo causal* da seleção natural – o traço está

presente na população por ter sido selecionado para tanto. Mas seria esta uma explicação *genuinamenteleológica* da existência do traço? Certamente, se por explicação teleológica entendermos a explicação da existência de uma entidade (e.g., traço ou organismo) em termos dos efeitos que a mesma acarreta (GARSON, 2016, p. 36). Assim, se explicamos que seres humanos possuem rins porque estes filtram o sangue, estamos explicando a presença dos rins em termos do efeito que estes acarretam, i.e., a filtragem do sangue.

Asegunda virtude é que a teoria etiológica explica à perfeição a distinção entre função e acidente. Por que o coração tem a função de bombear sangue e não de fazer o som característico de sua batida já que tanto o bombeamento quanto o som da batida são efeitos do coração? O que ocorre é que este foi *selecionado para* bombear sangue, enquanto o som da batida é apenas um *efeito colateral* do efeito para o qual o coração foi selecionado – bombear sangue (SOBER, 1984, pp. 99-100). Auxiliar a respiração é a função do meu nariz, não segurar óculos (embora este seja um efeito benéfico do nariz), porque os ancestrais do meu nariz foram selecionados para auxiliar a respiração, não para segurar óculos. Este último efeito é apenas um feliz acidente.

Finalmente, a terceira virtude é que a teoria etiológica explica o mau funcionamento dos traços, i.e., como traços podem ter uma dada função, mas falharem em exercê-la. Como um rim de um paciente pode ter a função de filtrar o sangue, se falha sistematicamente, a ponto do paciente necessitar de hemodiálise? Rins foram selecionados para filtrar o sangue, mas devido a uma doença ou má formação, podem falhar em fazer aquilo para o qual seus ancestrais foram selecionados. Como a teoria etiológica sustenta que é a história de seleção que determina a função de um traço, não suas capacidades atuais, é plenamente possível que o traço seja presentemente incapaz de fazer algo ainda que seus ancestrais tenham sido selecionados para fazerem isto.

À luz de tudo isto, não surpreende que a teoria etiológica seja atualmente a mais popular entre as teorias das funções biológicas. Porém, como sói acontecer com teorias filosóficas, grandes bônus vêm acompanhados de grandes ônus. A teoria etiológica enfrenta diversas dificuldades; no que se segue apresento duas das objeções mais fundamentais. A primeira objeção é que cientistas já atribuíam funções a traços biológicos *muito antes* do desenvolvimento da teoria da evolução pela seleção natural. Por exemplo, já no século XVII, William Harvey descobriu que a função do coração é bombear

sangue e essa atribuição funcional não foi feita à luz de qualquer processo de seleção. Mas se tal atribuição foi feita alheia a qualquer processo de seleção, como é possível que a função do coração seja bombear sangue, em virtude de seus ancestrais terem sido selecionado para tanto? De fato, o próprio Harvey julgava que os corações não foram sujeitos a qualquer processo de evolução pela seleção natural, posto que foram frutos da criação divina. Os cientistas usavam o conceito de função biológica muito antes de usarem o conceito de seleção natural, de modo que funções biológicas nada têm a ver com seleção natural. Em suma, a teoria etiológica não captura adequadamente a maneira como os cientistas usam o conceito de função.

A resposta etiológica mais influente é que esta objeção depende fundamentalmente da suposição de que teorias da função biológica devem consistir em *análises conceituais* da função biológica. Isto é, uma análise do que os biólogos significam em suas práticas científicas pelo conceito de função biológica, que almeja descrever como os biólogos usam e aplicam tal conceito. Contudo a abordagem etiológica rejeita justamente essa suposição, defendendo que teorias da função biológica devem desenvolver *definições teóricas* de funções biológicas (MILLIKAN, 1989, p. 291). O seguinte exemplo ilustra bem esta distinção entre análise conceitual e definições teóricas. A definição teórica de água como H₂O, ou ouro como o elemento de número 79, não é uma descrição de como as pessoas usam o conceito de ouro ou de água. Antes, almeja revelar a natureza última do ouro e da água, independentemente de como as pessoas aplicam os conceitos de ouro e água. Assim como não faz sentido objetar contra a definição teórica de água como H₂O, recorrendo à maneira como as pessoas usam o conceito de água, também não faz sentido objetar contra a definição teórica de função biológica como efeito selecionado, recorrendo à maneira como os biólogos usam o conceito de função. Contudo o debate quanto aos objetivos de uma teoria da função biológica – se devem desenvolver análises conceituais ou definições teóricas – é longo e complexo, estando para além dos propósitos da presente introdução.

A segunda objeção baseia-se em *contraexemplos* à teoria etiológica (casos reais ou hipotéticos) que costumam vir em duas categorias. Na primeira, estão aqueles itens que a teoria etiológica implica que são destituídos de função, por não terem sido sujeitos ao processo seletivo apropriado, mas que intuitivamente parecem ter uma função (RUSE, 1973). Isto é, casos que

contrariam a tese etiológica de que a seleção é um requisito necessário para ter função: o item parece ter função, mas não foi selecionado para tal. Considere uma mutação aleatória que se mostrou adaptativamente benéfica como o primeiro ancestral das zebras atuais a ter listras que detinha moscas. Uma vez que este primeiro caso não tinha uma história seletiva prévia na qual houve seleção de zebras para ter listras que detivessem moscas, segue-se que tais listras eram destituídas de função biológica. Contudo, é bastante intuitivo que as listras desta primeira zebra ancestral já tinham a função de deter moscas, o que contraria o veredito etiológico.

A *segunda categoria* de contraexemplos vai no sentido contrário: o item não tem a função, embora tenha sido selecionado para tal. Aqui estão os itens que a teoria etiológica implica ter uma função biológica, mas que intuitivamente são destituídos desta função. Nesta categoria, os casos paradigmáticos são itens não biológicos que foram sujeitos a um processo de seleção natural mas que intuitivamente não possuem função: populações de cristais de argila satisfazem os três critérios para seleção natural – variedade, reprodução diferenciada e hereditariedade – mas ninguém diria que cristais de argila têm função biológica (BEDAU, 1991). Assim como ocorreu com a primeira objeção à teoria etiológica, essa segunda objeção também suscitou um longo e complexo debate entre filósofos da biologia quanto ao seu sucesso (para uma avaliação crítica destas e outras objeções, cf. GARSON, 2016, pp. 45-55).

À luz destas dificuldades, muitos filósofos rejeitam a teoria etiológica e propuseram outras teorias. A principal rival da teoria etiológica foi primeiramente desenvolvida por Robert Cummins, a *teoria do papel causal* ou *sistêmica* da função biológica (CUMMINS, 1975). Ela não considera que o papel explicativo de atribuições funcionais na biologia consista em explicar a presença de traços nos organismos, mas em explicar o *funcionamento* de sistemas ou processos biológicos. O contraste com a abordagem etiológica é cristalino.

O sistema circulatório circula o sangue pelo corpo, transportando nutrientes, oxigênio e outras substâncias. Mas qual a contribuição do coração para o sistema circulatório? Cummins propõe que o papel explicativo da atribuição funcional de bombear sangue ao coração é justamente explicar como o coração *causalmente contribui* para o sistema circulatório circular e transportar sangue. Para determinar a contribuição de um componente para o sistema do qual este faz parte, é preciso fazer uma *análise funcional* deste sistema. Primeiro

se decompõe o sistema circulatório em seus diversos componentes: coração, veias, artérias, sangue etc. Em seguida, descreve-se a capacidade de cada componente: o coração bombeia o sangue, as veias e artérias transportam o sangue etc. É à luz desta decomposição que se descreve como o coração contribui para o sistema circulatório, a saber, bombeando o sangue a ser circulado. Note que só há circulação adequada de sangue quando o coração o bombeia. A análise funcional mostra como as subcapacidades de cada componente são organizadas de modo a contribuir para a capacidade final do sistema circulatório como um todo – circular e transportar o sangue.

Função Sistêmica: a função de cada traço do sistema biológico é a sua contribuição causal específica para a operação do sistema.

A função do componente é determinada por seu *papel causal* para o funcionamento do sistema do qual faz parte.

Enquanto a teoria etiológica sustenta que atribuições funcionais dão origem a *explicações teleológicas* para a *presença* de traços, a teoria do papel causal sustenta que atribuições funcionais dão origem a explicações *não teleológicas* de como os traços *funcionam* nos sistemas biológicos dos quais fazem parte. Por trás disso, está a rejeição por Cummins da própria teleologia biológica: atribuições e explicações funcionais na biologia não têm qualquer conotação teleológica. De acordo com a teoria do papel causal, atribuições funcionais são relevantes nas investigações de como traços e sistemas funcionam. Tais investigações procuram descrever a *interação causal* dos componentes do sistema, bem como entre cada componente e o sistema como um todo. Elas procuram determinar a contribuição causal de cada componente para o funcionamento do sistema. É no contexto dessas investigações que a atribuição funcional exerce o seu papel explicativo. A *grande virtude* da teoria do papel causal é fundamentar como atribuições funcionais têm um poder explicativo nas explicações de como os traços e sistemas biológicos funcionam *no presente*. Ela deixa de lado a história de seleção de traços e sistemas biológicos que constitui o foco da teoria etiológica (OKASHA, 2019, pp. 27-45). Contudo a teoria do papel causal sofre também diversos problemas que põem em xeque sua abordagem sistêmica da função biológica. No que se segue, apresentarei um dos mais fundamentais.

Karen Neander objeta que a teoria do papel causal é incapaz de fazer a necessária distinção entre o bom e mau funcionamento de traços biológicos (NEANDER, 1991; 2012). Se a função consiste em uma capacidade efetivamente já possuída por um componente do sistema (o bombeamento de sangue pelo coração) que contribui causalmente para a capacidade final do sistema (a circulação de sangue), então é impossível tal componente estar em mau funcionamento. A possibilidade do componente falhar em exercer sua função é simplesmente excluída pela teoria do papel causal. Porém só há função biológica, se houver possibilidade de mau funcionamento e essa teoria não abre espaço para tal possibilidade. A teoria do papel causal implica que o papel causal do traço é um dos papéis causais *efetivamente* já exercidos por este traço. Contudo a função de um traço não é necessariamente uma das capacidades que o traço já executa; é preciso distinguir quando há bom e mau funcionamento.¹⁰⁵

Uma maneira talvez mais clara de apresentar este problema é em termos da noção de disposição. O sal tem a disposição de se dissolver na água – o sal se dissolve quando uma condição é satisfeita, i.e., a presença de água. De acordo com a teoria do papel causal, a função de um traço é uma *disposição* deste traço. Isto é, o traço tem a função fazer F quando tem a disposição de fazer F sob tais-e-tais condições. Assumindo isto, se o coração se torna incapaz de bombear sangue, devido a uma doença, segue-se que ele deixa de ter a disposição de bombear sangue e, portanto, deixa de ter esta função. A teoria do papel causal não implica que o coração passa a estar em mau funcionamento, mas que ele *perde* a função de bombear sangue que possuía antes da doença. A possibilidade de o coração estar em mau funcionamento é, portanto, excluída pela teoria. O problema de se a teoria do papel causal exclui ou não a possibilidade de mau funcionamento é um dos principais focos da controvérsia entre os seus defensores e detratores.

O debate entre as teorias etiológica e sistêmica está no centro do

¹⁰⁵ “[Na morte devido ao câncer] conforme os tumores se espalham, substituem os tecidos saudáveis e a integridade funcional progressivamente deteriora. Se um tumor pressiona uma artéria no cérebro, este é um papel causal efetivo que o tumor possui neste processo patológico [...]. Contudo [ao contrário do que implica a teoria do papel causal] este não é a função biológica dos tumores. Tumores simplesmente não possuem função biológica [*proper function*].” (NEANDER, 1991, p. 181).

debate filosófico quanto à teleologia e à função biológica. Embora essas sejam as principais teorias, há outras disponíveis no mercado, como a teoria da propensão (BIGELOW & PARGETTER, 1987). Diante destes embates, tem-se tornado cada vez mais comum filósofos adotarem um *pluralismo funcional* quanto a teorias da função biológica: não há apenas *uma* teoria correta, mas *várias*. Pluralistas funcionais procuram compatibilizar certas teorias da função biológica e desenvolver critérios para quando e como uma teoria deve ser aplicada. Talvez haja diferentes sentidos da noção de função na biologia.

Ao longo desta seção, frequentemente assumi que explicações funcionais são explicações teleológicas e que estas últimas seriam indispensáveis na biologia. Contudo diversos filósofos *rejeitam* esta suposição. Como vimos, Cummins defende que explicações funcionais não são teleológicas, e muitos defendem que a biologia é destituída de qualquer teleologia. Como se pode ver, há um longo caminho para que o problema da teleologia na biologia seja finalmente solucionado.

4. Espécie

A unidade taxonômica básica da classificação biológica é a espécie. Trata-se de um dos pilares da nossa própria compreensão dos seres vivos, desempenhando um papel central em diversas áreas da biologia. Mas o que são espécies biológicas? Biólogos e filósofos sempre se fizeram esta pergunta. Nas palavras do famoso biólogo Ernest Mayr, “provavelmente não há nenhum outro conceito em biologia que tenha permanecido tão consistentemente controverso como o conceito de espécie” (MAYR, 1982, p. 251). O problema da espécie pode ser dividido em dois. O primeiro é o *problema da individuação* das espécies: o que *individua* as espécies, i.e., o que distingue uma espécie da outra? O que diferencia o mico-leão-dourado, o coqueiro e o bem-te-vi de modo a que pertençam a diferentes espécies? O segundo é o *problema do status ontológico* das espécies: qual o status ontológico das espécies? Seria a divisão dos organismos em espécies um aspecto *objetivo* e fundamental da realidade, i.e., independente do biólogo que desenvolva qualquer classificação dos organismos, ou seriam as espécies meras *projeções* do biólogo sobre a realidade? O objetivo desta seção é avaliar o problema da espécie a partir de sua divisão nos problemas da individuação e do status ontológico.

A biologia classifica os mais diversos organismos em espécies. São espécies nós, os seres humanos (*Homo sapiens*), os ursos pandas (*Ailuropoda melanoleuca*), a mandioca (*Manihot esculenta*), os micro-organismos responsáveis pela doença de Chagas (*Trypanosoma cruzi*) etc. A biologia se propõe a classificar todos os organismos em um dado táxon (unidade taxonômica), sendo os principais táxons, em ordens crescente, os seguintes: espécie, gênero, família, ordem, classe, filo e reino.¹⁰⁶ Primeiramente cada organismo é atribuído a uma espécie. Então um conjunto de espécies são reunidas formando um gênero, um conjunto de gêneros formam uma família e assim sucessivamente, até um conjunto de filos formar um reino. Por exemplo, a espécie *Homo sapiens* pertence ao gênero *Homo*, à família dos hominídeos, à ordem dos primatas, à classe dos mamíferos, ao filo dos cordados, e finalmente ao reino animal. A espécie é o táxon básico, toda a classificação dos seres vivos depende fundamentalmente do conceito de espécie. As espécies desempenham um papel fundamental na biologia, de modo que se não soubermos o que é uma espécie, o nosso edifício hierárquico corre risco de desabar, o que mostra a importância do problema da espécie na biologia e filosofia da biologia.

As ciências naturais classificam seus objetos de investigação. A física classifica os estados da matéria em diversos tipos, sendo os mais famosos os estados líquido, sólido e gasoso. A química classifica os compostos em orgânicos e inorgânicos. Com a biologia não poderia ser diferente: a taxonomia é o ramo da biologia responsável pela classificação dos organismos. Mas por que os cientistas têm a necessidade de classificar seus objetos de investigação? Há diversas razões para tanto, como a de *carregarinformações* a respeito dos objetos reunidos em uma dada classificação. Por exemplo, a classificação de dois organismos numa mesma espécie implica que estes organismos têm diversas características em comum. As classificações científicas reúnem objetos que são semelhantes em diversos aspectos, a fim de descrever como os mesmos se comportam e manifestam.

Contudo nem toda classificação reflete uma distinção *real e objetiva* do mundo. Afinal, os objetos podem ser classificados das mais diversas maneiras,

¹⁰⁶Acima do táxon dos reinos biológicos há ainda o táxon dos domínios biológicos (*Eukarya*, *Bacteria* e *Archaea*). O naturalista Carlos Lineu é em geral considerado o pai da taxonomia moderna (RICHARDS, 2016).

independentemente de como o mundo esteja objetivamente organizado. Isto significa que os objetos podem ser classificados das maneiras mais *arbitrárias* possíveis. A questão que se coloca é a seguinte: a partir de qual *critério* devemos avaliar classificações alternativas dos objetos de um domínio? Há uma maneira correta de classificá-los ou todas as classificações são em última instância igualmente corretas?

No caso das espécies, um observador qualquer pode classificar todos os organismos em duas categorias excludentes, a saber, aqueles que gosta e aqueles que desgosta. Os chimpanzés são classificados numa categoria juntamente com os mosquitos da dengue, enquanto os gorilas são classificados na categoria oposta simplesmente porque tal observador desgosta dos primeiros, mas gosta dos últimos, por mais que chimpanzés e gorilas sejam organismos muito mais semelhantes entre si do que chimpanzés e mosquitos da dengue. Mas a qual critério devemos recorrer para avaliar as classificações dos organismos? Note que a taxonomia biológica classifica chimpanzés e gorilas na mesma família dos homínídeos, enquanto chimpanzés e mosquitos-da-dengue não são sequer classificados no mesmo filo – os primeiros pertencem ao filo dos cordados enquanto os mosquitos da dengue pertencem ao filo dos artrópodes. Em virtude de quê devemos preferir a classificação biológica em detrimento da classificação do observador acima? Sustentar que esta última deve ser rejeitada por ser arbitrária não soluciona o problema posto, de que o que buscamos é justamente a *razão* pela qual a classificação deste observador é arbitrária em contraste com a classificação biológica.

Iniciemos pelo problema da individuação: o que distingue uma espécie da outra? Sua solução exige uma definição do que seja uma espécie. Contudo não há consenso entre os biólogos e filósofos quanto ao que seja uma espécie. Definições rivais, também chamadas “conceitos de espécie”, resultam em classificações distintas das espécies: certos organismos são classificados por uma definição como pertencentes a uma espécie, enquanto que por outra são classificados como pertencentes a outra espécie. Isto é, definições rivais traçam diferentes limites para distinguir uma espécie da outra – em uma definição, uma certa variação em um grupo de organismos conta como uma variação *no interior* de uma mesma espécie, ao passo que em outra definição esta mesma variação conta como uma variação *entre* espécies. No que se segue, avalio algumas das principais definições de espécie.

Talvez a definição mais imediata de espécie seja o *conceito fenético de espécie* (CFE): uma espécie é um grupo de organismos *intrinsecamente similares* em um dado aspecto – morfologia, genética, comportamento etc. (SOKAL & SNEATH, 1963). Gorilas e mosquitos-da-dengue não são da mesma espécie em virtude de não serem semelhantes à luz de um destes aspectos – e.g., são organismos com morfologia completamente diferentes. O CFE recorre à semelhança de organismos de uma mesma espécie e a transforma na própria definição do que é uma espécie.

Contudo, o CFE é uma definição muito problemática por diversas razões (STERELNY & GRIFFITHS, 1999, pp. 184-6). Em primeiro lugar, a viabilidade do CFE enquanto definição de espécie depende de uma *medição objetiva* da semelhança dos organismos. O problema é que há uma quantidade colossal de diferentes métodos para calcular a similaridade, que, por vezes, produzem diferentes medições. Mas então como decidir de maneira não arbitrária e não *ad hoc* entre todos estes métodos? Isto parece ser impossível. Em segundo lugar, não é verdade que os organismos que pertencem a uma mesma espécie sejam *sempre* similares uns aos outros. Por exemplo, machos, fêmeas e jovens são dessemelhantes em diversos aspectos como morfologia e comportamento. Há espécies, como a dos iguanas, nas quais os machos têm estratégias de procriação muito diferentes uns dos outros e, portanto, são organismos bem diferentes uns dos outros. Ainda mais grave, diferentes organismos de uma mesma espécie podem se assemelhar mais a organismos de outra espécie do que se assemelham aos organismos de sua própria espécie. À luz dessas e de outras razões, o CFE é uma definição atualmente amplamente rejeitada.

A definição de espécie mais influente é o *conceito biológico de espécie* (CBE), que tem em Mayr o seu principal proponente: uma espécie é um grupo de organismos *reprodutivamente isolado* de outros grupos de organismos (MAYR, 1942). Isto significa que organismos pertencentes a uma mesma espécie são capazes de *se reproduzirem entre si*, enquanto organismos pertencentes a espécies diferentes são reprodutivamente isolados. Chimpanzés e mosquitos da dengue não são organismos da mesma espécie, por serem reprodutivamente isolados, portanto incapazes de procriação. Chimpanzés e gorilas, apesar de muito mais semelhantes, também não pertencem a mesma espécie por também não serem capazes de procriar. Já os cães pastor alemão e doberman

pertencem à mesma espécie já que são capazes de se reproduzirem entre si, ainda que pertençam a raças distintas. A ideia por trás do CBE é que as semelhanças e dessemelhanças entre os organismos que resultam na especiação têm origem na *restrição do fluxo de genes*. A semelhança entre os chimpanzés surgiu e persistiu ao longo da história evolutiva devido ao isolamento reprodutivo dos chimpanzés. Por não procriarem com outros organismos, a aparência e comportamento dos chimpanzés persistem. Os genes mutantes dos chimpanzés ficam restritos a este grupo e não se espalham no *pool genético* (a totalidade de genes) de outros organismos, já que eles não se reproduzem entre si, não havendo livre fluxo genético entre os chimpanzés e os demais organismos. Note como o CBE *explica* a frequente similaridade dos organismos de uma espécie, ao apelar para o isolamento reprodutivo, em vez de definir espécie em termos de similaridade como o CFE erroneamente o faz.

O CBE foi um marco na compreensão biológica da especiação. Mas como costuma ocorrer com definições em geral, a viabilidade do CBE é ameaçada em várias frentes. Em primeiro lugar, ele, de partida, exclui organismos assexuados de serem membros de qualquer espécie, posto que não pertencem a grupos reprodutivamente isolados. De fato, não pertencem a qualquer grupo reprodutivo – isolado ou não. Talvez aqui a saída seja encarar o CBE não como uma definição geral de espécie, mas como restrita a certos tipos de organismos – aqueles com reprodução sexuada. Em segundo lugar, o próprio Mayr notou que o CBE não pode ser aplicado a grupos de organismos geograficamente isolados (MAYR, 1942, pp. 120-1). Afinal, isolamento geográfico não implica necessariamente isolamento reprodutivo e, a menos que dois grupos de organismos estejam em contato, não podemos verificar se são capazes de se reproduzir ou não. Por fim, há o problema do hibridismo – a reprodução entre organismos de espécies diferentes. No caso das plantas, é bastante comum haver procriação entre organismos que pertencem a espécies diferentes, embora também haja hibridismo entre animais (e.g., cacatua-rosa e galah). O CBE traça aqui uma classificação equivocada posto que estas plantas e animais pertencem a espécies diferentes apesar de não serem reprodutivamente isoladas (OKASHA, 2019, p. 70; STERELNY & GRIFFITHS, 1999, pp. 189).

À luz de todos os problemas enfrentados pelo CBE, dezenas de outras definições de espécie foram propostas. Uma das mais influentes é o *conceito ecológico de espécie* (CEE) que define espécie em termos de nichos ecológicos.

Contudo está para além dos propósitos desta introdução avaliar a viabilidade destas outras definições ou de uma potencial compatibilidade entre as mesmas. A solução para o problema da individuação das espécies está longe de ser desenvolvida. Passemos ao problema do status.

Suponha que de fato exista uma classificação correta das espécies. Qual é então o *status ontológico* das espécies? A ontologia é a área da filosofia que investiga aquilo que há, ou seja, aquilo que existe. O problema do status ontológico, portanto, consiste no problema sobre a maneira como a espécie existe, i.e., o seu modo de ser. Seriam as espécies entidade reais e objetivas? A classificação dos organismos em espécies independe dos interesses e das arbitrariedades do biólogo que assim os classifica, tendo as espécies, portanto, uma constituição independente de qualquer observador?

A resposta mais tradicional ao longo da história da filosofia para o problema do status ontológico é o *essencialismo*. Espécies possuem essências que determinam sua realidade e objetividade, independentes de qualquer observador. Mas o que é uma essência? Esta é uma questão complexa, mas à primeira vista a essência de um objeto é aquilo sem o qual este não seria aquilo que ele é. Por exemplo, é supostamente uma propriedade essencial de um ser humano qualquer ser um animal, mas não a propriedade de estar localizado à minha direita. Afinal, deixando de ser um animal, também deixa de ser humano, ao passo que continuaria a ser humano, se estivesse à minha esquerda.

O essencialismo de espécie é um essencialismo de tipo: a espécie é um *tipo natural com essência*. Como vimos anteriormente, há múltiplas maneiras de classificar objetos no mundo em tipos, mas apenas algumas destas classificações são *naturais* na medida em que correspondem a divisões objetivas da realidade, não refletindo os interesses ou arbitrariedades de qualquer observador. Classificações naturais são aquelas que cortam a realidade nas juntas, para usar a famosa metáfora de Platão no *Fedro* (PLATÃO, 2016, 265e). É em virtude de sua essência que o organismo é membro da espécie a qual ele efetivamente pertence: é essencial para este coqueiro pertencer à espécie dos coqueiros. A tese da espécie como um tipo natural vêm acompanhada de alguns princípios, destaquemos dois destes. O primeiro é que todos os organismos de uma espécie têm uma essência comum que não é compartilhada pelos organismos que não pertencem a esta espécie – uma certa propriedade *necessária e suficiente* para um organismo pertencer a certa espécie. Assim, há

uma propriedade essencial exclusivamente compartilhada pelos coqueiros. O segundo princípio é que esta essência comum é responsável pelas características tipicamente associadas aos organismos da espécie. Assim, é a essência dos coqueiros que fundamenta o fato de eles terem o côco como fruto.

Contudo, o essencialismo quanto à espécie tornou-se amplamente rejeitado após o estabelecimento da teoria da evolução pela seleção natural. O essencialismo foi uma tese ontológica bastante plausível num mundo pré-darwiniano, no qual Deus criou cada uma das espécies, que seriam eternas e imutáveis. Deus dotou cada espécie com uma essência intrínseca, que fundamentaria as características típicas de seus organismos constitutivos. A principal objeção ao essencialismo é a aparente ausência de propriedades exclusivamente compartilhadas pelos membros da espécie. Não parece haver nenhuma propriedade intrínseca como morfologia, comportamento ou genética que seja essencial. O principal problema é que há *suficiente variação* morfológica, comportamental e genética entre os membros da espécie, que impede que haja uma propriedade compartilhada por todos os seus organismos constitutivos e por nenhum organismo de outra espécie. Note que a evolução ocorre através de uma contínua mudança na base genética dos membros da espécie, seja por reprodução sexual ou por mutação genética. Ou seja, as forças evolutivas atuam continuamente contra a exclusividade de qualquer traço pelos membros da espécie. A evolução mostra que as espécies estão em contínua mudança, de modo que, ao longo da história, as espécies não apenas se modificam, como podem ser extintas. Por fim, ainda que por hipótese, em um intervalo de tempo na história evolutiva de uma espécie, haja uma propriedade compartilhada tão somente por seus membros, estes organismos podem passar por mudanças que façam com que nem todos compartilhem mais esta propriedade. Estas dificuldades levaram muitos filósofos a rejeitarem o essencialismo de espécie, sustentando que “o essencialismo quanto às espécies é hoje um assunto encerrado” (SOBER, 1980, p. 353).

Desde a derrocada do essencialismo biológico no século XX, diversas teorias não essencialistas foram propostas para tentar dar conta do status ontológico das espécies. Contudo, nas últimas décadas, houve o ressurgimento do essencialismo biológico em uma nova versão que pode ser batizada de *essencialismo relacional* (para não ser confundido com o essencialismo tradicional) segundo o qual espécies seriam *tipos naturais históricos*. Isto é,

seriam tipos naturais com *essências históricas*, não com *essências intrínsecas*, tal como proposto pelo essencialismo tradicional. Antes, seriam *essências extrínsecas ou relacionais*, posto que constituídas pela história evolutiva dos organismos (GRIFFITHS, 1999; LaPORTE, 2004).

Propriedades podem ser classificadas como intrínsecas ou extrínsecas. Propriedades intrínsecas são aquelas que o objeto possui tão somente em virtude de si mesmo, propriedades cuja instanciação depende exclusivamente dele próprio. Por outro lado, propriedades extrínsecas são aquelas que o objeto possui em virtude também de algo para além dele mesmo, propriedades cuja instanciação não dependem exclusivamente dele próprio. Dito de outra forma, propriedades extrínsecas exigem referência a algo externo, ao contrário de propriedades intrínsecas. Considere uma caixa de alumínio. Sua massa é uma propriedade intrínseca, posto que depende exclusivamente de sua constituição interna, ao passo que seu peso é uma propriedade extrínseca, por depender também da gravidade do ambiente na qual a caixa está inserida (esta possui um peso na Terra e outro em Marte). Eis o porquê de propriedades extrínsecas serem também chamadas de *propriedades relacionais*.

Pois bem, essencialistas relacionais defendem que o essencialismo tradicional cai por terra, por propor que as propriedades essenciais que determinam o pertencimento a uma espécie sejam intrínsecas (e.g., propriedades genéticas ou morfológicas). Eis que surgem todas aquelas dificuldades anteriormente elencadas, que acabam por minar o essencialismo tradicional. Contudo essas dificuldades não se levantam, desde que tais propriedades essenciais sejam relacionais. Simplesmente não há nenhuma propriedade *intrínseca* que seja exclusivamente compartilhada por todos os membros de uma espécie, mas há uma propriedade *relacional*.

O essencialismo relacional define espécie em termos de *linhagem evolutiva*. A ideia é que um organismo pertence a uma dada espécie devido a seus genitores pertencerem a esta espécie. A *descendência* é a propriedade em virtude da qual o organismo pertence a esta ou aquela espécie. Uma vez que a descendência é determinada pela linhagem evolutiva da qual o organismo faz parte, segue-se que a descendência é uma propriedade histórica e, portanto, relacional. A essência da espécie nada mais é que a sua posição na árvore evolutiva. As relações evolutivas do organismo são necessárias e suficientes para este pertencer à espécie que efetivamente pertence. Por exemplo,

Aristóteles é um *Homo sapiens* não em virtude de qualquer propriedade morfológica ou genética que ele possa vir a ter, mas por seus pais serem *Homo sapiens*, seus avós serem *Homo sapiens*, etc. Em suma, por Aristóteles pertencer à linhagem evolutiva dos *Homo sapiens*. Segundo Paul Griffiths, nada que não tenha a origem histórica comum dos membros de uma espécie pode pertencer à espécie.¹⁰⁷ O essencialismo relacional defende que é esta propriedade histórica que explica os traços frequentemente atribuídos a cada espécie, como morfologia, comportamento, genótipo etc. Contudo, tais traços não são propriedades essenciais, posto que não são necessários e suficientes para o organismo ser desta ou daquela espécie. Apenas a origem histórica é essencial.

Poderia o essencialismo relacional finalmente solucionar o problema do status ontológico? Um dos problemas que ele enfrenta é justamente o seu status enquanto uma teoria *essencialista* das espécies: poderiam propriedades históricas relacionais serem propriedades genuinamente *essenciais*? Ou apenas propriedades intrínsecas poderiam ser genuinamente essenciais? (GODFREY-SMITH, 2014, p. 112-3). É inegável que a concepção de essencialidade se aplica mais naturalmente às propriedades intrínsecas, afinal essencial é aquilo em virtude do qual o objeto é aquilo que ele é. Não é a toa que tradicionalmente as propriedades essenciais foram vistas como intrínsecas. Por exemplo, historicamente essências foram vistas como propriedades que se manifestam nas características superficiais dos objetos que as possuem. A natureza interna de um objeto pode ser vista como se expressando na maneira como o objeto se comporta (e.g., devido à sua constituição interna, o dióxido de carbono se apresenta no estado gasoso quando em temperatura ambiente). Mas não está claro que propriedades extrínsecas possam assim se expressar no comportamento do objeto. A objeção então para o essencialismo relacional é que propriedades históricas relacionais como pertencer a uma linhagem evolutiva não podem ser propriedades essenciais do organismo, justamente por serem extrínsecas ao organismo. Segundo Godfrey-Smith, “usar a palavra ‘essência’

¹⁰⁷“nada que não compartilhe a origem histórica do tipo pode ser um membro deste tipo. Embora Lilith pudesse não ter sido um gato doméstico, como um gato doméstico, ela é necessariamente um membro do nexa genealógico entre o evento de especiação no qual este táxon se originou e o evento de especiação ou extinção na qual ele irá deixar de existir.” (GRIFFITHS, 1999, p. 219).

para propriedades extrínsecas é agarrar-se a uma palavra carregada, quando muito de seu significado original foi retirado” (GODFREY-SMITH, 2014, p. 113).

O essencialismo relacional também foi alvo de várias outras objeções, incluindo aquelas advindas de seu próprio campo essencialista (DEVITT, 2008). Contudo está para além do escopo desta introdução avaliá-las. Concluo apenas sublinhando que o essencialismo biológico, nas suas mais diversas versões, é, ainda hoje, uma das principais posições no debate quanto ao status ontológico das espécies. A notícia sobre a sua morte foi bastante exagerada, como diria Mark Twain.

5. Conclusão

A filosofia da biologia é uma das áreas mais fascinantes e frutíferas da filosofia contemporânea. Meu objetivo nesta introdução foi apresentar ao leitor alguns de seus principais tópicos. Todavia, devido ao limite de espaço, tive de deixar de fora outros tópicos tão ou mais relevantes como aqueles aqui apresentados. Decidi perder na quantidade de problemas apresentados para ganhar na profundidade da apresentação dos problemas aqui tratados. Entre os problemas que foram deixados de fora, são dignos de nota aqueles quanto à unidade da seleção, ao adaptacionismo, ao reducionismo na biologia e às leis biológicas. Ao leitor interessado, há boas introduções em língua portuguesa a estes e outros problemas da filosofia da biologia, cf. CHEDIK, 2008; ABRANTES, 2018.

Questões de revisão

1. Quais são os três princípios da evolução pela seleção natural? Aplique-os para explicar a evolução de algum traço fenotípico.
2. Explique a diferença entre as teorias etiológicas e sistêmicas da função biológica. Quais as principais vantagens e desvantagens destas teorias?
3. Explique a diferença entre propriedades essenciais e inessenciais. Posteriormente, explique a diferença entre propriedades intrínsecas e extrínsecas. Baseando-se nestas duas distinções, explique a diferença entre o essencialismo de espécie tradicional e o essencialismo relacional.

Questões de discussão

4. A teoria da evolução é incompatível com uma interpretação literal da bíblia segundo a qual Deus criou separadamente espécies imutáveis. Porém a teoria da evolução não parece ser, em si mesma, incompatível com a hipótese de um Deus que criou o universo. Por que não?
5. De uma perspectiva pluralista, como as teorias etiológicas e sistêmicas poderiam ser compatíveis? Talvez por diferentes propósitos teóricos exigirem diferentes concepções de função biológica. Desenvolva.
6. Suponha que em última instância o essencialismo de espécie fracasse, tanto na sua versão tradicional como na sua versão relacional. Como você conceberia uma teoria não essencialista da espécie? Quais seriam suas vantagens e desvantagens?

Sugestões de leitura

Evolução

Uma ótima introdução à teoria da evolução pela seleção natural para o grande público seja o clássico *O Gene Egoísta* de Richard Dawkins (DAWKINS 1976 [2007],). Outra ótima introdução que foca nas evidências científicas que apoiam a teoria da evolução é *Por que a evolução é uma verdade*, de Jerry Coyne (COYNE 2009 [2014]). A melhor introdução à teoria da evolução que foca em seus aspectos filosoficamente mais relevantes é *Sex and Death: an Introduction to Philosophy of Biology*, de Kim Sterelny e Paul Griffiths (STERELNY & GRIFFITHS, 1999). A controvérsia entre o criacionismo e a teoria da evolução é longa e complexa, não se resumindo tão somente ao debate entre evolucionistas e intérpretes literais da bíblia. Também envolve debates quanto à própria relação entre ciência e religião, bem como quanto às implicações sócio-políticas desta controvérsia. Para os aspectos filosóficos desta controvérsia que se encontram na interseção entre as filosofias da religião, da ciência, da biologia e da política, cf. RUSE, 1988, 2018; PENNOCK, 2007; ROBERTS, 2010. Quanto à teoria do design inteligente, cf. NUMBERS, 2010; RUSE, 2009.

Função

Para uma introdução ao problema da teleologia e da função biológica, cf. OKASHA, 2019; NEANDER, 2012. Para a melhor avaliação crítica das teorias das funções biológicas, cf. GARSON, 2016. Em língua portuguesa, cf. CHEDIAK, 2018. Para o debate entre aqueles que defendem e rejeitam a teleologia na biologia, cf. WALSH, 2008; ARIEW, 2007; NEANDER, 2018.

Espécie

Para uma introdução ao problema da espécie, cf. GODFREY-SMITH, 2014; RICHARDS, 2016. Para uma avaliação dos problemas da individuação e do status ontológico, cf. STERELNY & GRIFFITHS, 1999; RICHARDS, 2010; ERESHEFSKY, 2017. Para outras teorias da espécie, cf. BOYD, 1999; MILLIKAN, 1999; GHISELIN, 1974; HULL, 1978.

Referências

- ABRANCHES, P. C. (org). 2018 *Filosofia da biologia*. Editora do PPGFIL - UFRRJ. Seropédica, RJ. Disponível em <https://repositorio.unb.br/bitstream/10482/32184/1/LIVRO_FilosofiaBiologia.pdf>. Acesso em: 11 de ago. de 2020.
- ARIEW, A. 2007. *Teleology*. In “The Cambridge companion to philosophy of biology” (HULL & RUSE, org.), Cambridge University Press, Cambridge, pp. 1-21.
- BEDAU, 1991. *Can biological teleology be naturalized?*. The Journal of Philosophy, 88, 11, pp. 647-55.
- BIGELOW, J. & PARGETTER, R. 1987. *Functions*. Journal of Philosophy, 84, 4, pp. 181–196.
- BOYD, R. 1999. *Homeostasis, Species, and Higher Taxa*. In “Species: New Interdisciplinary Studies” (WILSON, ed.), MIT Press, Cambridge, MA, pp. 141-185.
- CHEDIAK, K. 2008. *Filosofia da biologia*. Jorge Zahar Editor. Rio de Janeiro.
- _____. 2018. *Funções e explicações funcionais em biologia*. In “Filosofia da Biologia” (ABRANCHES, org.), Editoria do PPGFIL - UFRRJ, Seropédica,

- RJ, pp. 103-121.
- COYNE, J. A. 2009 [2014]. *Por que a evolução é uma verdade*. JSN Editora. São Paulo.
- CUMMINS, R. 1975. *Functional analysis*. The Journal of Philosophy, 72, pp. 741-765.
- DARWIN, C. & WALLACE, A. R. 1858. *On the tendencies of species to form varieties; and on the perpetuation of varieties and species by natural means of selection*. Journal of the Proceedings of the Linnean Society of London. Zoology, 3, 9, pp. 45–62.
- DARWIN, 1859 [2018]. *A Origem das espécies por meio de seleção natural*. Ubu Editora. São Paulo.
- DAWKINS, 1976 [2007]. *O Gene egoísta*. Companhia das Letras. São Paulo.
- _____. 1986 [2001]. *O Relojoeiro cego*. Companhia das Letras. São Paulo.
- DOBZHANSKY, T. 1973. *Nothing in biology makes sense except in the light of evolution*. The American Biology Teacher. 35, 3, pp. 125-129.
- DEVITT, M. 2008. *Resurrecting biological essentialism*. Philosophy of Science, 75, 3, pp. 344-382.
- ERESHEFSKY, M. 2017. *Species*. The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2017/entries/species/>>. Acesso em: 11 de ago. de 2020.
- GARSON, J. 2016. *A Critical overview of biological functions*. Springer International Publishing. Cham, Switzerland.
- GHISELIN, M. 1974. *A Radical solution to the species problem*. Systematic Zoology, 23, pp. 536–544.
- GILDENHUYS, P. 2019. *Natural selection*. The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2019/entries/natural-selection/>>. Acesso em: 11 de ago. de 2020.
- GRIFFITHS, P. 1999. *Squaring the circle: natural kinds with historical essences*. In “Species: new interdisciplinary studies” (WILSON, ed.), MIT Press, Cambridge, MA, pp. 209-228.
- GODFREY-SMITH, P. 2014. *Philosophy of biology*. Princeton University Press. Princeton.
- HULL, D. 1978. *A Matter of individuality*. Philosophy of Science, 45, pp. 335–360.

- JACOBS, J. 2020. *Naturalism*. Internet Encyclopedia of Philosophy. URL = <<https://iep.utm.edu/naturali/>>. Acesso em: 11 de ago. de 2020.
- LEWONTIN, R. C. 1970. *The Units of selection*. Annual Review of Ecology and Systematics, 1, pp. 1-18.
- _____. 1978. *Adaptation*. Scientific American, 239, pp. 212-28.
- LaPORTE, J. 2004. *Natural kinds and conceptual change*. Cambridge University Press, Cambridge.
- MAYR, E. 1942. *Systematics and the origins of species*. Columbia University Press. New York.
- _____. 1982. *The growth of biological thought*. Harvard University Press. Cambridge, MA.
- MILLIKAN, R. G. 1989. *In defense of proper functions*. Philosophy of Science, 56, pp. 288-302.
- _____. 1999. *Historical kinds and the "special sciences"*. Philosophical Studies, 95: 45-65.
- NEANDER, K. 1991. *Functions as selected effects: the conceptual analyst's defense*. Philosophy of Science, 58, pp. 168-184.
- _____. 2012. *Biological function*. Routledge Encyclopedia of Philosophy, Taylor and Francis. URL = <<https://www.rep.routledge.com/articles/thematic/biological-function/v-1>>. Acesso em: 11 de ago. de 2020.
- _____. 2018. *Does biology need teleology?*. In The "Routledge handbook of philosophy and evolution"(JOYCE, ed.), Routledge, New York and London, pp. 64-76.
- NUMBERS, R. 2010. *Criacionismo científico e design inteligente*. In "Ciência e Religião" (HARRISON, P., ed). Ideias e Letras. São Paulo., pp. 165-190.
- PENNOCK, R. T. 2007. *Biology and religion*. In "The Cambridge companion to philosophy of biology" (HULL & RUSE, eds.), Cambridge University Press, Cambridge, pp. 410-428.
- ROBERTS, J. H. 2010. *Reações religiosas a Darwin*. In "Ciência e Religião" (HARRISON, ed.), Ideias e Letras, São Paulo, pp. 109-38.
- OKASHA, S. 2019. *Philosophy of biology*. Oxford University Press. Oxford.
- ODENBAUGH, J. & GRIFFITHS, P. 2020. *Philosophy of biology*. The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2020/entries/biology->

- philosophy/>. Acesso em: 11 de ago. de 2020.
- PLATÃO. 2016. *Fedro*. Companhia das Letras. São Paulo.
- PENNOCK, R. 2007. *Biology and religion*. In *The Cambridge companion to philosophy of biology* (HULL & RUSE, eds.), Cambridge University Press, Cambridge, pp. 410-428.
- RICHARDS, R. 2010. *The Species problem: a philosophical analysis*. Cambridge University Press, Cambridge.
- _____. 2016. *Biological classification: a philosophical introduction*. Cambridge University Press, Cambridge.
- ROSENBERG, A. & McSHEA, D. W. 2008. *Philosophy of biology: a contemporary introduction*. Routledge, Nova York & Abingdon.
- RUSE, M. 1973. *A reply to Wright's analysis of functional statements*. *Philosophy of Science*, 40, pp. 277–280.
- _____. 1988. *But is it science? The Philosophical question in the creation/evolution controversy*. Prometheus, Buffalo, N.Y.
- _____. 2009. *Myth 23 – that “intelligent design” represents a scientific challenge to evolution*. In “Galileo goes to jail and other myths about science and religion” (NUMBERS, ed.), Harvard University Press, Harvard, pp. 206-14.
- _____. 2018. *Creationism*. *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2018/entries/creationism/>>. Acesso em: 11 de ago. de 2020.
- SOBER, E. 1980. *Evolution, population thinking, and essentialism*. *Philosophy of Science*, 47, 3, pp. 350-83.
- _____. 1984. *The Nature of selection: evolutionary theory in philosophical focus*. MIT Press. Cambridge, MA.
- SOKAL, R. R. & SNEATH, P. H. 1963. *Principles of numerical taxonomy*. W. H. Freeman and Company. San Francisco.
- SOUZA FILHO, S. 2013. *Seguir regras e naturalismo semântico*. Dissertação (Mestrado em Filosofia) – Instituto de Filosofia e Ciências Sociais, Universidade Federal do Rio de Janeiro. Rio de Janeiro. Disponível em: <<https://ppglm.files.wordpress.com/2008/12/dissertacao-ppglm-sergio-souza.pdf>>. Acesso em: 11 de ago. de 2020.
- STERELNY, K. & GRIFFITHS, P. E. 1999. *Sex and death: an introduction to philosophy of biology*. University of Chicago Press. Chicago and London.

- SPINOZA, B. 2009 [1677]. *Ética*. Autêntica. Belo Horizonte.
- TABERY, J. 2019. *Genetics*. The Stanford Encyclopedia of Philosophy (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2020/entries/genetics/>>. Acesso em: 11 de ago. de 2020.
- WALSH, D. M. 2008. *Teleology*. In “The Oxford Handbook of Philosophy of Biology” (RUSE, org.), Oxford University Press, Oxford, pp. 113-137.
- WRIGHT, L. 1973. *Functions*. *Philosophical Review*, 82, 139–168.

9

Filosofia da Linguagem

Sagid Salles

1. Introdução

A filosofia da linguagem é a parte da filosofia que lida com problemas filosóficos que dizem respeito à linguagem. A linguagem esconde – e por vezes nem esconde tão bem – uma infinidade de problemas filosóficos. Os principais problemas surgem em torno de dois conceitos fundamentais: *referência* e *significado*. Grosso modo, usamos o primeiro conceito para falar da suposta conexão entre expressões linguísticas e o mundo, e o segundo para falar daquilo que comunicamos por meio destas expressões. Mesmo o mais inocente uso da linguagem pode esconder problemas filosóficos relacionados à referência e ao significado. Considere, por exemplo, a frase abaixo.

1. Aristóteles é sábio.

Há incontáveis objetos no universo, mas ao proferir (1) estou a falar de apenas um deles: Aristóteles. De algum modo, meu uso do nome “Aristóteles” me permitiu selecionar um único indivíduo dentre todos os outros, para então poder dizer algo sobre ele. Este é apenas um exemplo de como a nossa linguagem pode corresponder ao mundo ou a partes dele, e em filosofia queremos compreender conexões deste tipo. Mas este nem de longe é o único mistério levantado por (1). Sabemos que (1) é dotada de significado, que não se trata de mero ruído. Ao ler (1), você está certo de que *compreendeu* o que eu disse. Em filosofia, também estamos interessados em explicar o fenômeno do significado e o fenômeno correlacionado da compreensão linguística. Imagino que não seja necessário perder tempo explicando a importância deste empreendimento. Afinal, se há alguma coisa da qual (quase) ninguém duvida, é

a importância da linguagem natural – aquela linguagem que usamos em nosso cotidiano. Não fosse pelo fato de tudo correr relativamente bem com a linguagem – em particular no que diz respeito aos fenômenos da referência, significado e compreensão – o leitor sequer conseguiria ler este texto. Aliás, esse nem de longe seria seu maior problema. Tente passar um único dia sem recurso à linguagem e terá uma pequena dimensão da importância desta ferramenta em nossas vidas. Para entender como esta ferramenta funciona, precisamos lidar com os problemas filosóficos que ela origina, em particular aqueles que dizem respeito aos conceitos de *referência* e *significado*.

Para o bem ou para o mal, há mais problemas em torno destes conceitos do que eu poderia sequer sonhar em abordar em um texto introdutório. Alguma seleção é inevitável, e não hesitei em deixar com que minhas próprias preferências me influenciassem neste aspecto. Conforme ficará claro mais tarde, dei maior foco aos chamados “problema fundacional da referência” e “problema descritivo do significado”, e escolhi os nomes próprios como exemplo de discussão acerca destes problemas. Neste sentido, o presente capítulo é uma introdução pessoal ao tema, e o caminho trilhado ao longo dele é parcialmente baseado no caminho que eu próprio trilhei em meu contato com a filosofia da linguagem. Ao longo do caminho, teremos a oportunidade de considerar muitos conceitos fundamentais, mas é importante ressaltar que muita coisa foi deixada de fora. A despeito disto, penso que a discussão aqui apresentada será suficiente para dar uma ideia do tipo de problemas com os quais a filosofia da linguagem lida, e de como nós filósofos lidamos com os mesmos. Se tiver sucesso, a exposição será suficientemente introdutória para permitir que leitores pouco informados, porém interessados, consigam despertar a curiosidade sobre os problemas e soluções discutidos. A partir daí, caberá ao leitor decidir se deseja se aprofundar no tema.

2. O que é uma teoria da referência?

As expressões “Aristóteles” e “sábio” parecem estabelecer algum tipo especial de conexão com o mundo. Pretensamente, ambas correspondem a algo: “Aristóteles” corresponde a Aristóteles, e “sábio” corresponde aos sábios. O mesmo vale para nomes próprios e adjetivos em geral, e também para expressões de muitas outras categorias. Para citar apenas um exemplo, a

descrição definida “a pessoa mais rica do planeta” pretensamente corresponde ao indivíduo mais rico do planeta. Já expressões como “aliás”, “portanto”, “prol”, etc. não correspondem a qualquer coisa, e sequer há a pretensão de que o façam. A função de “Aristóteles” em (1) é selecionar um objeto do mundo, mas a função de “aliás” em “Aliás, Aristóteles é sábio” não está relacionada a um objeto ou conjunto de objetos.

As expressões do primeiro grupo levantam pelo menos dois tipos de problemas que não são levantados pelas do segundo. O primeiro é determinar a quais itens do mundo elas correspondem, quando de fato correspondem a algo. Em alguns casos, a solução para este problema parece ser muito fácil. Nomes próprios, por exemplo, correspondem a objetos particulares, que podem ser pessoas, animais de estimação, locais, construções etc. O nome “Aristóteles” corresponde à pessoa Aristóteles, e o nome “Torre Eiffel” corresponde àquela famosa torre em Paris. Mas as coisas nem sempre são tão simples. Mesmo no caso dos nomes próprios, filósofos nem sempre estão em paz sobre o que corresponde ao quê. Há amplo desacordo sobre se nomes ficcionais, como “Sherlock Holmes” e “Clark Kent”, se referem a algum tipo de *objeto ficcional*, e neste caso o que exatamente seriam estes objetos.¹⁰⁸ Com relação a outros tipos de expressões, como os adjetivos, o desacordo talvez seja ainda maior. Há dúvidas sobre se expressões como “sábio” correspondem a conjuntos de objetos atuais, conjuntos de objetos possíveis, funções, propriedades etc. Este tipo de debate está ligado a discussões metafísicas sobre o que existe e quais são as categorias fundamentais da realidade. O segundo problema é o de determinar como exatamente a relação de correspondência é estabelecida. Ora, sabemos que as palavras não passam a corresponder ao que correspondem por mero acaso, magia ou algo do tipo. Tem de haver algum mecanismo cujo bom ou mal funcionamento determina o sucesso ou insucesso da correspondência com o

¹⁰⁸Uma razão que pode levar alguém a pensar que nomes ficcionais correspondem a algo é o *paradoxo das existenciais negativas*. Inspirados por Meinong (1904), muitos tentam resolver este paradoxo, através da postulação de um novo tipo de objetos, que serviria como referente dos nomes ficcionais. Outros, inspirados na teoria das descrições de Russell (1905), ou na teoria do faz de conta desenvolvida por Kendall Walton (1990), acreditam poder resolver o problema sem a postulação de um novo tipo de objetos. Para uma discussão deste paradoxo, com foco no último tipo de solução, veja meu Salles (2015).

mundo. Queremos saber que tipo de mecanismo é este e como exatamente ele funciona. Este problema é mais restrito à filosofia da linguagem, e menos dependente de discussões metafísicas, do que o anterior.

Em acordo com uma distinção similar feita por Robert Stalnaker (1997, p. 535), podemos classificar estes dois tipos de problemas respectivamente como o *problema descritivo* e o *problema fundacional da referência*. Uma *teoria da referência* tem como objetivo resolver estes problemas. Talvez porque o segundo esteja mais intimamente relacionado à filosofia da linguagem, a expressão “teoria da referência” é frequentemente restrita a ele. Na ausência de qualquer indicação para o contrário, é assim que entenderei o termo daqui por diante.

3. Termos singulares e termos gerais (e quantificadores)

Vimos anteriormente que “Aristóteles” e “sábio” pertencem a um grupo especial de expressões que pretensamente correspondem a itens do mundo, e que uma teoria da referência pretende explicar como esta conexão ocorre. Agora veremos que estas são expressões de tipos diferentes, e que pode ser prudente investigar o mecanismo referencial de cada uma delas separadamente. De fato, esta é a prática mais comum entre filósofos. Isto também nos dará a oportunidade de introduzir dois conceitos importantes para a filosofia da linguagem: os conceitos de *termo singular* e *termo geral*. Comece por considerar as seguintes frases.

1. *Aristóteles* é sábio.
2. *Isto* é uma cadeira. [Apontando para uma cadeira específica].
3. *Aqui* é *Ouro Preto*.

Em (1) “Aristóteles” corresponde a Aristóteles, e “sábio” corresponde a cada (humano) sábio, enquanto em (2) “isto” corresponde a uma cadeira contextualmente relevante e “cadeira” corresponde a cada cadeira, e em (3) “aqui” corresponde a um lugar contextualmente relevante e “Ouro Preto” a Ouro Preto.

Repare que algumas expressões nas frases acima estão em itálico, enquanto outras não. Embora ambas pretensamente correspondam a objetos do mundo, existe uma diferença entre aquelas que estão em itálico e aquelas que

não estão. As primeiras satisfazem o que Sainsbury (2006) chama de “condição de no máximo um”: se elas correspondem a algum objeto, então correspondem a no máximo um objeto. Por outras palavras, tais expressões são usadas de forma a pretensamente corresponder a um e no máximo um objeto. Este é claramente o caso do nome “Aristóteles”, por exemplo. Para que uma frase como (1) seja verdadeira, tem de haver um e no máximo um indivíduo selecionado por “Aristóteles”. Compare a situação com a da palavra “sábio”. Poderia ser o caso que esta palavra correspondesse a apenas um indivíduo do mundo. Mas não há qualquer exigência de que seja assim. Para que (1) seja verdadeira, não é necessário que haja apenas um indivíduo ao qual a palavra “sábio” corresponda.

A condição *deno máximo um* se aplica apenas às expressões em itálico nas frases acima. Podemos usar isto para marcar a diferença entre dois grupos de expressões: *termos singulares* e *termos gerais ou predicados*. Tanto termos singulares como termos gerais são expressões que pretensamente correspondem a objetos, mas apenas os primeiros devem satisfazer a condição de no máximo um. Isto pode ser tornado um pouco mais claro, se pensarmos nas condições de verdade das frases acima. (1) é verdadeira se, e somente se, o objeto selecionado por “Aristóteles” pertence ao conjunto de objetos aos quais “sábio” corresponde. (2) é verdadeira se, e somente se, o objeto contextualmente selecionado por “isto” pertence ao conjunto de objetos aos quais “cadeira” corresponde. (3) é verdadeira se, e somente se, o objeto contextualmente selecionado por “aqui” é o mesmo objeto selecionado por “Ouro Preto”. Os nomes “Aristóteles” e “Ouro Preto”, e os indexicais “isto” e “aqui”, são termos singulares, enquanto o adjetivo “sábio” e o substantivo “cadeira” são termos gerais.

É importante notar que esta é apenas uma caracterização inicial, que de modo algum esgota o problema.¹⁰⁹ Por exemplo, tanto quanto a condição de no máximo um nos conta, não podemos simplesmente classificar o nome próprio “Aristóteles” como um termo singular. Isto ocorre porque nem todas as ocorrências de “Aristóteles” satisfazem esta condição. Considere a frase (4).

4. Há mais de um Aristóteles nesta sala.

¹⁰⁹ Para uma discussão interessante, veja Strawson (1959, seção 6.1)

O nome “Aristóteles”, em (4), não é usado sob a restrição da condição de no máximo um. Ao contrário, ele é usado como um termo geral. Uma solução provisória para o problema seria falar de expressões de nossa linguagem comum *sendo usadas como termos singulares e como termos gerais*. Strawson (1952, p. 145) corretamente sugere que a distinção entre termos singulares e termos gerais pode ser colocada em termos de uma distinção entre usos de palavras, entre as diferentes funções para as quais palavras podem ser usadas. Tome apenas aquelas expressões que pretensamente envolvem uma conexão especial com o mundo, e assuma que somente os termos singulares e gerais fazem parte daquele grupo. Quando uma delas é usada para selecionar um e no máximo um objeto, então é usada como um termo singular; quando é usada sem esta restrição, é usada como um termo geral. Isto ainda nos permitiria dizer que nomes próprios e adjetivos são respectivamente termos singulares e gerais, mas apenas no sentido de que eles são geralmente usados como tais.

Embora esta solução provisória nos permita caminhar com alguma segurança, ela não é completamente satisfatória. Talvez o principal problema aqui diga respeito ao caso das descrições definidas. Descrições definidas são expressões da forma “o/a F”, como “o autor da República”, “a rainha da Inglaterra” etc. Tal como cotidianamente usadas, as descrições satisfazem a condição de no máximo um. Mas muitos filósofos pensam que isto não basta para dizermos que elas são (ou estão sendo usadas como) termos singulares (nos contextos relevantes). Geralmente, estes filósofos estão munidos de definições mais precisas e rigorosas dos termos singulares e gerais. Estas definições podem envolver outros aspectos que não a condição de no máximo um. Pode-se, por exemplo, delimitar as categorias em termos do modo como contribuem para o significado ou para as condições de verdade das frases completas que os contêm, ou então em termos do papel inferencial que têm em certos contextos de uso. Sobre a primeira estratégia, veja Evans (1982, sec. 2.3) e Sainsbury (2006); sobre a segunda, veja Imaguire (2007).

Sem entrar em maiores detalhes sobre a distinção, há outro grupo de expressões que parece se conectar com o mundo de forma especial. Considere os seguintes exemplos.

5. *Alguém* foi reprovado.
6. *Todos* foram reprovados.

À primeira vista, “alguém” e “todos” correspondem a itens do mundo. Afinal, “alguém” corresponde a alguém, e “todos” corresponde a todos. Mas nem sempre as coisas são o que parecem. Estes termos são chamados de “quantificadores”. Na análise da lógica de predicados de primeira ordem, quantificadores não selecionam objetos ou conjuntos de objetos. Ao invés disto, eles dizem respeito à frequência com que certos termos gerais se aplicam aos objetos de um domínio relevante. Suponha, por exemplo, que (5) e (6) sejam proferidas por um professor para falar dos alunos de uma disciplina específica. Neste caso, o domínio relevante é o domínio de alunos da disciplina em questão. Grosso modo, o que (5) diz é que o termo geral “reprovado” se aplica a pelo menos um objeto deste domínio, isto é, a pelo menos um aluno daquela disciplina. Por sua vez, (6) diz que o predicado “reprovado” se aplica a todos os objetos deste domínio, a todos os alunos da disciplina relevante. Enquanto termos singulares pretensamente selecionam um objeto particular, e termos gerais se aplicam a objetos sem a restrição da condição de no máximo um, quantificadores nos contam com que frequência um termo geral se aplica aos objetos de um dado domínio. Se isto estiver correto, então quantificadores não selecionam objetos, ou pelo menos não no sentido em que termos singulares e gerais o fazem. Não é claro, portanto, que estas expressões devam ser incluídas no grupo especial mencionado acima.

Agora repare nas duas perguntas que seguem:

- i. O que faz com que um termo singular corresponda ao objeto que ele seleciona?
- ii. O que faz com que um termo geral corresponda aos objetos aos quais ele corresponde?

Certamente deve haver algum mecanismo garantindo o sucesso ou insucesso dos termos singulares em selecionar objetos particulares. Do mesmo modo, deve haver algum mecanismo garantindo o sucesso ou insucesso dos termos gerais em corresponder aos objetos aos quais correspondem. Mas é bem provável que haja diferenças significativas entre os dois casos. Por outras palavras, o problema fundacional da referência dos termos singulares pode envolver aspectos que não estão presentes no problema fundacional da referência dos termos gerais, e vice-versa. Uma teoria que responde à pergunta

(i) é uma *teoria da referência dos termos singulares*, enquanto uma teoria que responde à pergunta (ii) é uma *teoria da referência dos termos gerais*. Alguns filósofos, como Michael Devitt (1981, p. 9), reservam a palavra “aplicação” para o segundo caso. Neste contexto, podemos fazer a distinção entre uma *teoria da referência dos termos singulares* e uma *teoria da aplicação dos termos gerais*.

4. Termos singulares

Mesmo o projeto de uma teoria da referência singular pode ser demasiado ambicioso, dado que diferentes tipos de expressões podem envolver diferentes mecanismos de referência singular. Por exemplo, o mecanismo referencial de indexicais puros, verdadeiros demonstrativos e nomes próprios não parece funcionar do mesmo modo.

David Kaplan (1977, p. 491) argumentou plausivelmente a favor de que existe uma distinção entre *indexicais puros* e *verdadeiros demonstrativos*. Os primeiros são associados a um significado linguístico que determina um referente para cada contexto de uso, de modo que não são necessárias quaisquer intenções ulteriores por parte do falante. Um possível exemplo de indexical puro é o indexical “eu”. O significado linguístico de “eu” é uma regra que determina que o referente desta expressão será quem quer que tenha a proferido no contexto de uso. Por exemplo, se João diz “eu sou legal”, então João é o referente de “eu”, mas se Maria diz isto, então ela é o referente de “eu”. Tão logo alguém profira esta expressão em um contexto de uso, será o referente da mesma, independentemente de quaisquer intenções ulteriores. O significado linguístico de um verdadeiro demonstrativo, por outro lado, não determina um referente para cada contexto de uso. Para que um demonstrativo tenha um referente em um contexto, será ainda necessário um ato de demonstração ou uma intenção direcionadora (KAPLAN, 1989, p. 582). Um exemplo de verdadeiro demonstrativo é a expressão “este”. Suponha que João diga “este quadro é soberbo”. Para que ele tenha sucesso em se referir a um quadro particular através de seu proferimento, é necessário que ele realize um ato de demonstração (como apontar o dedo para o quadro relevante) ou pelo menos tenha a intenção de falar de um quadro específico. Se Kaplan estiver certo – e penso que ele de fato está – então a explicação da referência dos indexicais puros deve envolver o apelo aos significados linguísticos mencionados acima,

enquanto a explicação da referência dos verdadeiros demonstrativos deve apelar a atos de demonstração ou intenções direcionadoras. Nenhum destes elementos deve ser incluído em uma explicação da referência dos nomes próprios. Intuitivamente, o mecanismo referencial dos nomes não envolve o tipo de significado linguístico relevante para os indexicais puros, nem atos de demonstração ou intenções direcionadoras.

Dentre as expressões que são geralmente usadas como indexicais puros provavelmente estão “eu”, “hoje”, “amanhã”, “atual”, “presente” etc. Dentre aquelas que são geralmente usadas como verdadeiros demonstrativos provavelmente estão “isto”, “este”, “aquele”, “ele”, “ela” etc. Já nomes próprios incluem expressões como “Aristóteles”, “Maria”, “Sócrates” etc. Se as observações acima estão corretas, então temos motivação para estudar o mecanismo referencial de cada uma dessas categorias de expressões independentemente. Ao invés de perseguir o projeto de uma teoria da referência singular, podemos preferir o projeto de uma *teoria da referência dos indexicais puros* ou uma *teoria da referência dos verdadeiros demonstrativos* ou uma *teoria da referência dos nomes próprios*.

Existe uma lição a aprender do que foi dito até aqui. A filosofia da linguagem pode ter objetivos grandiosos, como o de entender a relação entre a linguagem e o mundo. Mas não é possível alcançar estes objetivos sem atenção às peculiaridades de diferentes tipos de expressões. Esta é parte da razão pela qual filósofos da linguagem gastam tanto tempo discutindo detalhes à primeira vista irrelevantes. A opção de ignorar estes detalhes em nossa investigação da linguagem seria metodologicamente desastrosa.

5. O que é uma teoria do significado?

Até agora nos concentramos no fato de que “Aristóteles” e “sábio” pretensamente correspondem a objetos do mundo. Mas este não é o único fato interessante sobre estas expressões. Outro fato interessante é que elas são dotadas de significado. Aliás, o mesmo se aplica à frase (1) como um todo. Compare-a, por exemplo, com (7).

1. Aristóteles é sábio.
7. Konokaieve blu blu.

Ao contrário de (1), (7) é apenas uma sequência aleatória de símbolos que nada significam. Isto ressalta dois tipos de problema. O primeiro é qual o significado de uma expressão linguística, seja ela uma frase completa ou alguma expressão contida na frase (como um termo singular ou um termo geral). Conforme veremos, a tarefa de determinar o significado de uma expressão é frequentemente complicada. O segundo é em virtude de quais fatos acerca de nós ou da nossa prática linguística as expressões têm o significado que elas têm. Obviamente, expressões não possuem seu significado em virtude do mero acaso. Tem de haver um conjunto de fatos em virtude dos quais (1) tem significado, enquanto (7) não; e em virtude dos quais (1) significa exatamente o que significa, ao invés de qualquer outra coisa. Filósofos estão interessados em descobrir quais são estes fatos. Em acordo, uma vez mais, com uma distinção similar feita por Stalnaker (1997, p. 535), podemos classificar estes dois tipos de problemas respectivamente como o *problema descritivo* e o *problema fundacional do significado*. Uma *teoria do significado* tem como objetivo resolver estes problemas.

No restante deste artigo, me concentro no problema descritivo do significado. Como veremos, há relações interessantes entre as respostas a este problema e as respostas ao problema fundacional da referência. É importante notar que, a despeito de ser ignorado aqui, o problema fundacional do significado têm recebido ampla atenção dos filósofos contemporâneo e, atualmente, ocupa posição central na filosofia da linguagem. Uma instigante apresentação introdutória do tema pode ser encontrada em Lycan (2000).

6. Composicionalidade

Compare (1) com (8) abaixo.

8. Marcos é sábio.

Ambas são dotadas de significado, mas significados diferentes. O que explica esta diferença? A resposta imediata é que a diferença se deve ao fato de elas conterem expressões diferentes na posição de sujeito. O mesmo ocorre com (1) e (9), que contém expressões diferentes na posição de predicado.

1. Aristóteles é sábio.
9. Aristóteles é barbudo.

Mas então considere (10) e (11) abaixo.

10. A neve é branca.
11. Snow is white.

Nenhuma expressão contida em (10) está presente em (11), e vice-versa. Ainda assim, (10) e (11) têm o mesmo significado. Isto mostra que do fato de duas frases conterem expressões diferentes não se segue que tenham significados diferentes. De forma similar, do fato de duas frases conterem exatamente as mesmas expressões não se segue que tenham o mesmo significado, como podemos facilmente notar através de (12) e (13).

12. João chegou ao banco [instituição financeira].
13. João chegou ao banco [local para sentar].

Apesar de (12) e (13) conterem exatamente as mesmas expressões, elas não significam a mesma coisa. Isto porque a palavra “banco” é usada com um sentido diferente em cada caso. Em resumo, a identidade ou diferença entre os significados de duas (ou mais) frases não depende de se elas são compostas ou não das mesmas expressões. A identidade ou diferença depende de se as expressões componentes têm ou não o mesmo significado. (10) e (11) têm o mesmo significado em parte porque suas expressões componentes significam a mesma coisa, e (12) e (13) têm significados diferentes porque pelo menos uma de suas expressões significam coisas diferentes. Mas isto ainda não é o bastante, como podemos notar através de (14) e (15).

14. João ama Maria.
15. Maria ama João.

(14) e (15) não apenas são compostas das mesmas expressões, estas expressões também significam a mesma coisa em cada caso. Ainda assim, elas possuem significados diferentes. O problema agora reside no modo como as

expressões estão organizadas, na estrutura das frases. Embora as duas frases acima contenham as mesmas expressões com os mesmos significados, elas são estruturadas de modo diferente. Como sabemos, faz toda diferença neste caso se “João” ocorre antes ou depois de “Maria”. Moral da história: o significado de uma frase parece depender pelo menos do significado de suas expressões componentes e de sua estrutura. Daí o *princípio da composicionalidade*, que afirma que o significado de uma frase é determinado pelo significado de suas partes e pelo modo como estas estão estruturadas. Filósofos frequentemente falam do significado de uma frase como a *proposição* ou o *Pensamento* expressado por ela. Neste contexto, podemos dizer que o princípio da composicionalidade afirma que a proposição ou Pensamento expressado por uma frase é determinado pelo significado de suas partes e por sua estrutura.

Além de ser muito intuitivo, o princípio da composicionalidade tem uma importância especial em discussões sobre o conceito de *compreensão linguística*. Em particular, muitos acreditam que este princípio explica a nossa incrível capacidade de compreender frases novas, com as quais nunca tivemos contato. Estou certo de que será a primeira vez que o leitor se depara com a seguinte frase: Em 2014, marcianos verdes interferiram na concentração dos jogadores da seleção brasileira e fizeram com que a última perdesse de 7 x 1 para a Alemanha, em pleno Mineirão! A despeito de ser seu primeiro contato com esta frase, imagino que não tenha tido qualquer dificuldade em compreendê-la. Grosso modo, isto ocorre porque você conhece o significado de cada expressão componente, assim como as regras que regem suas combinações. Faça o teste por si mesmo: substitua alguma expressão da frase acima por uma expressão cujo significado lhe seja estranho, ou embaralhe as expressões de forma a resultar em uma estrutura desconhecida. Em seguida, pergunte-se: sou capaz de entender perfeitamente a frase resultante? A resposta será “não”.

Razões deste tipo fazem com que o princípio da composicionalidade seja bastante popular entre filósofos da linguagem – o que não significa que seja uma unanimidade.¹¹⁰ Neste contexto, não é surpreendente que muitos filósofos pensem que o significado de uma expressão subfrásica deve ser entendido em termos do modo como ela contribui para o significado das frases completas que

¹¹⁰Uma ótima apresentação introdutória da discussão acerca da composicionalidade é Szabó (2020).

a contêm. Por outras palavras, para resolver o problema descritivo do significado para expressões como “Aristóteles” e “sábio”, por exemplo, devemos entender como estas expressões contribuem para a proposição ou para o pensamento expresso pelas frases completas nas quais ocorrem. Se isto estiver correto, então o princípio da composicionalidade tem uma importância fundamental na discussão sobre o problema descritivo do significado. Seja como for, isto será assumido daqui por diante.

Anteriormente, vimos que há motivação para investigarmos de forma independente o problema fundacional da referência para diferentes tipos de expressões. O mesmo se aplica ao problema descritivo do significado. A princípio, é bastante plausível que expressões de diferentes categorias contribuam de forma diferente para a proposição expressa pelas frases completas que as contêm. Por exemplo, termos singulares e gerais não parecem contribuir da mesma forma. Neste contexto, ao invés de perseguir o projeto de uma teoria do significado, podemos preferir o projeto de uma *teoria do significado dos termos gerais* ou de uma *teoria do significado dos termos singulares*. No último caso, se quisermos ser ainda mais específicos, podemos optar pelo projeto de uma *teoria do significado dos indexicais puros*, *teoria do significado dos verdadeiros demonstrativos* ou *teoria do significado dos nomes próprios*.

7. Conteúdo x significado linguístico

Vimos que o significado de uma frase é a proposição expressa por ela, enquanto o significado de uma expressão subfrásica é entendido em termos da contribuição que ela faz para a proposição expressa pelas frases completas que a contêm. Seguindo Kaplan (1977, p. 500), podemos dizer que o significado de uma frase ou expressão, neste sentido, é o *conteúdo* da mesma. É importante não confundir conteúdo com o que foi anteriormente chamado de “significado linguístico”.

O significado linguístico de uma expressão – seja uma frase ou uma expressão subfrásica – é convencional e não varia de contexto para contexto. Além disto, embora o significado linguístico restrinja o conteúdo que uma expressão pode ter em um contexto, ele frequentemente não é rico o bastante para determinar o conteúdo propriamente dito (RECANATI, 2004, p. 5-6; CARSTON, 2002, p. 17-18). Por exemplo, o significado linguístico da frase “eu

sou filósofo” impede que esta frase expresse (literalmente) a proposição de que a neve é branca. Apesar disto, ele não é rico o bastante para determinar qual proposição ela expressa. Isto ocorre porque, dado o significado linguístico de uma das expressões componentes – o indexical “eu” – a frase “eu sou filósofo” pode expressar que Sagid é filósofo (quando proferida por mim), que Maria é filósofa (quando proferida por Maria), que o João é filósofo (quando proferida por João), e assim por diante.

A expressão “significado” é ambígua entre conteúdo e significado linguístico. Em qualquer que seja o sentido, a ideia de significado levanta tanto o problema descritivo como o problema fundacional. Por um lado, podemos perguntar qual é o significado linguístico ou qual é o conteúdo de uma expressão (problema descritivo). Por outro, podemos querer saber em virtude de quais fatos uma expressão tem determinado significado linguístico ou em virtude de quais fatos ela tem certo conteúdo (problema fundacional). Aliás, o princípio de composicionalidade pode ser sustentado para ambas as noções de significado. Assim como dissemos que o conteúdo de uma frase é determinado pelo conteúdo de suas partes mais o modo como estão estruturadas, alguém poderia dizer que o significado linguístico de uma frase é determinado pelo significado linguístico de suas partes mais o modo como estão estruturadas.

Daqui por diante, vou usar a expressão “significado” para falar exclusivamente do *conteúdo* de frases e expressões subfrásicas. Quando quiser falar do significado linguístico, deixarei isto explícito.

8. O que é dito x o que é implicitado¹¹¹

O conceito de *significado*, mesmo quando entendido no sentido de *conteúdo*, é ambíguo. Primeiro, pode-se entender o conteúdo de uma frase como *o que é dito* por ela. É neste sentido que o conteúdo é fortemente restringido pelo significado linguístico. Como vimos na seção anterior, embora o significado linguístico de “eu sou filósofo” não seja rico o bastante para determinar o que é dito, ele restringe o que pode ser dito por esta frase. Esta frase pode dizer que

¹¹¹“Implicitar” corresponde ao inglês “implicate”, introduzido por Grice em um de seus clássicos artigos (GRICE, 1967). Neste ponto, estou a seguir a tradução que utilizo do artigo.

Sagid é um filósofo, ou que João é um filósofo, ou... Todavia, ela não pode dizer que eu não sei resolver uma determinada equação matemática.

Existe, contudo, um sentido de conteúdo em que ele não está limitado pelo significado linguístico, ou pelo menos não de forma tão severa. Trata-se de um sentido no qual o conteúdo de um proferimento vai além do que é dito pelo mesmo. Suponha que um amigo me pergunte qual é o resultado de uma equação e, em resposta, eu lhe diga “Eu sou filósofo”. Neste caso, consegui de algum modo ir além do que meu proferimento diz, e transmiti a informação de que não sei resolver a equação proposta. Agora modifique um pouco o exemplo, e imagine que meu amigo me pergunte qual é a minha profissão. Então lhe respondo dizendo “Eu sou filósofo”, enquanto pisco descaradamente um dos olhos. Neste caso, embora meu proferimento diga que eu sou um filósofo, consegui de algum modo transmitir a informação de que *não* sou um filósofo. Por fim, suponha que meu amigo me pergunte se gosto de Platão e, em resposta, lhe diga “eu sou filósofo”. Neste caso, a despeito do que meu proferimento diz, consegui transmitir a informação de que gosto de Platão. Estes exemplos apontam para uma distinção entre *o que é dito* e *o que é implicitado*. Embora em cada um dos casos meu proferimento diga a mesma coisa – que sou um filósofo – ele implica diferentes coisas; ele implica respectivamente que não sei resolver a equação matemática proposta, que *não* sou um filósofo e que gosto de Platão.

Enquanto o que é dito é severamente restringido pelo significado linguístico, o que é implicitado não sofre este tipo de restrição. Um proferimento de “eu sou filósofo” pode implicar praticamente qualquer coisa, se as circunstâncias forem adequadas. “O que é implicitado é implicitado em virtude de uma inferência, e a cadeia de inferência pode (em princípio) ser tão longa e envolver tantas suposições de fundo quanto alguém desejar” (RECANATI, 2001, p. 76, tradução minha).

Este processo de inferência pode ocorrer de modos variados. No caso mais famoso, chamado por Grice (1967) de “implicaturas conversacionais”, o ouvinte consegue inferir o que é implicitado pelo falante a partir de seu conhecimento do que é dito pelo proferimento do falante, e da suposição de que o falante é cooperativo e respeita certas máximas conversacionais. De forma simplificada, no primeiro exemplo acima o ouvinte pode recorrer a um raciocínio do seguinte tipo para inferir que estou a comunicar que não sei resolver a

equação. (a) O proferimento do Sagid – “Eu sou filósofo” – diz que ele é um filósofo. (b) Ou o Sagid está violando a máxima conversacional de acordo com a qual devemos ser pertinentes, ou ele comunica que não sabe resolver o problema matemático. (c) Não há razão aparente para supor que ele esteja a violar a máxima em questão. (d) Sagid está comunicando que não sabe resolver o problema relevante.

Obviamente, filósofos estão interessados na compreensão de ambos os sentidos de *significado*; seja no problema de determinar o que é dito e o que é implicado por um proferimento (problema descritivo do significado), seja no problema de saber em virtude de quais fatos o que é dito/implicado é dito/implicado (problema fundacional do significado). Além disto, a distinção entre o que é dito e o que é implicado é com frequência relacionada com duas outras distinções importantes para a filosofia da linguagem: significado literal x significado não literal, e semântica x pragmática. Quanto ao primeiro caso, pode-se pensar que o significado literal corresponde ao que é dito, e o significado não literal ao que é implicado.¹¹² Quanto ao segundo, pode-se pensar que a semântica se ocupa do que é dito, enquanto a pragmática se ocupa do que é implicado.¹¹³ Não tratarei destes problemas aqui. Daqui por diante, quando falar em significado, estarei me referindo ao que é dito por um proferimento, e não ao que é implicado.

9. Referência direta e indireta

Não é incomum que as discussões sobre referência e significado apareçam misturadas. De fato, é possível que nossa resposta a uma classe de problemas dependa de nossa resposta à outra. Desde Frege (1892), há um conjunto de filósofos que pensam que o significado das expressões desempenha algum papel importante na determinação de seu referente. Mais especificamente,

¹¹²A despeito de possíveis diferenças na terminologia, parece-me que minimalistas semânticos como Bach (1994, 2006b) e Cappelen & Lepore (2005) aceitam um significado literal fortemente conectado ao que é dito, enquanto contextualistas como Recanati (2004) rejeitam a própria ideia de significado literal.

¹¹³Eu colocaria a distinção entre semântica x pragmática entre as mais obscuras da filosofia da linguagem. Um artigo iluminante sobre o tema é Bach (1997). Veja também Travis (1997).

o referente de uma expressão é pelo menos parcialmente determinado pelo significado da mesma. Esta é a tese da *referência indireta*. Aqueles que recusam isso defendem a tese da *referência direta*. Neste sentido, a última afirma que seja qual for o meio pelo qual o referente de uma expressão é determinado, o significado não desempenha qualquer papel nesta determinação. Por outras palavras, dizer que uma expressão se refere indiretamente a algo é dizer que sua referência é de algum modo mediada por seu significado, e dizer que uma expressão se refere diretamente é negar que haja tal mediação.

Frege pensava que a distinção entre significado e referente se aplicava não apenas a expressões subfrásicas, mas também a frases completas. Assim, uma frase pode ter tanto um significado quanto um referente. O significado de uma frase é, como dito, o Pensamento ou proposição expressa por ela. O seu referente seria seu valor de verdade: o Verdadeiro ou o Falso. Também neste caso, o referente seria pelo menos parcialmente determinado pelo significado. A tentativa de Frege de aplicar a distinção também às frases é controversa, e faremos melhor em deixá-la de lado.¹¹⁴

Uma vez que as teses da referência direta e indireta nos informam algo sobre como o referente de uma expressão linguística é determinado, elas dizem respeito ao problema fundacional da referência. Como veremos, no entanto, esta discussão está intimamente relacionada ao problema descritivo do significado. Isso ficará mais claro a partir da próxima seção, quando começo a aplicar os conceitos discutidos até aqui a uma discussão particular.

10. Referencialismo e nomes próprios

O *referencialismo* é uma resposta ao problema descritivo do significado, de acordo com a qual o significado de uma expressão é apenas seu referente. No caso de expressões subfrásicas, isto quer dizer que a única contribuição que estas expressões fazem para o significado das frases completas que as contêm é a introdução de um referente no discurso. Se expressões linguísticas não possuem qualquer significado para além de seu referente, então não é o significado que determina o referente. Dizer que o significado determina o

¹¹⁴Para uma breve porém clara apresentação da semântica fregeana veja Kemp (2013, cap. 2).

referente seria, neste caso, equivalente a dizer o referente determina a si mesmo; e isto é claramente insatisfatório. Por esta razão, o referencialismo implica a referência direta. Ou seja, trata-se de uma teoria (descritiva) do significado que implica a rejeição da teoria da referência indireta, que é parte de uma teoria (fundacional) da referência. Temos aqui um exemplo de como o problema descritivo do significado e o problema fundacional da referência se entrelaçam.

Infelizmente, o referencialismo tem um pequeno problema: ele é obviamente falso. Afinal, já vimos que nem todas as expressões dotadas de significado têm a função de introduzir referentes no discurso. A expressão “afinal”, que inicia a frase anterior, é um exemplo disto. Outros exemplos foram dados na segunda seção deste artigo. Não é por mero acidente que o referencialismo não desfruta de qualquer popularidade entre filósofos da linguagem.

As coisas ficam melhores quando abandonamos uma interpretação imperialista do referencialismo e o restringimos a categorias específicas de expressões. Se o referencialismo não parece correto com respeito à linguagem em geral, ele é bem plausível quando aplicado especificamente aos termos singulares, ou talvez a certa classe de termos singulares. Nomes próprios, por exemplo, de fato parecem meras etiquetas cuja única função é introduzir um objeto no discurso. A única função de “Aristóteles” em “Aristóteles é sábio” é introduzir o próprio Aristóteles no discurso, para que então possamos lhe atribuir a propriedade de ser sábio. Podemos exprimir isto dizendo que nomes simplesmente não possuem qualquer significado (HAACK, 1978, p. 94). Dado que nomes não possuem significado (seu significado é apenas seu referente), a relação de referência não pode ser mediada pelo significado do nome. A referência direta aparece uma vez mais como consequência.

A ideia de que o significado dos termos singulares, em particular dos nomes próprios, é apenas seu referente é muito popular entre filósofos. Mas não está isenta de problemas...

11. Frege e o ataque ao referencialismo

Frege (1982) é o autor do mais influente argumento contra o referencialismo. Considere as duas frases abaixo.

16. Héspero é Héspero.

17. Héspero é Eósforo.

De acordo com o princípio da composicionalidade, o significado de uma frase é determinado pelo significado de suas partes e por sua estrutura. Sabemos que (16) e (17) possuem a mesma estrutura. Atualmente, também sabemos que os nomes “Héspero” e “Eósforo” se referem ao mesmo objeto: o planeta Vênus. De acordo com o referencialismo, podemos concluir que estes nomes significam a mesma coisa. Resultado: (16) e (17) possuem o mesmo significado. Esta conclusão é problemática por pelo menos duas razões. Primeiro, (16) é uma identidade trivial conhecível *a priori*, enquanto (17) foi uma importante descoberta astronômica. Segundo, (16) e (17) têm *valores cognitivos* diferentes. Isto significa que um mesmo sujeito pode coerentemente sustentar *atitudes proposicionais* opostas com relação à proposição expressada por elas. Uma atitude proposicional é uma atitude que temos com respeito a uma proposição *p*, como *saber* que *p*, *acreditar* em *p*, *duvidar* de *p* etc. Por exemplo, é evidente que uma mesma pessoa poderia acreditar na proposição expressa por (16) e rejeitar (17), sem incorrer em qualquer tipo de incoerência. Em resumo, o referencialismo falha em explicar (a) por que (16), mas não (17), é uma identidade trivial, e (b) por que elas têm diferentes valores cognitivos.

Frege resolve o problema alegando que o significado de um nome próprio não é idêntico ao seu referente. O significado – ou *sentido* – de um nome é o modo de apresentação de seu referente. É possível que dois nomes diferentes, como “Héspero” e “Eósforo”, apresentem o mesmo objeto de modos diferentes. Isto explica (a) e (b) acima. (16) é uma identidade trivial, porque apresenta o mesmo objeto do mesmo modo duas vezes, enquanto (17) o apresenta de dois modos diferentes. A diferença de valor cognitivo se deve ao fato de que valor cognitivo não depende apenas do referente, mas também do sentido das expressões envolvidas. Compare com o valor de verdade (que Frege pensa ser o referente das frases). Ao substituirmos a segunda ocorrência de “Héspero” em (16) por “Eósforo”, não alteramos o valor de verdade da frase original. O resultado da substituição é (17), que tem exatamente o mesmo valor de verdade que (16): ambas são verdadeiras. O mesmo ocorreria se substituíssemos qualquer um dos nomes em (16) e (17) por “Vênus”, que se refere ao mesmo planeta. Em resumo, quando substituímos uma expressão

destas frases por alguma outra que tem o mesmo referente, o valor de verdade da frase não se altera.¹¹⁵ Em contrapartida, a substituição de “Héspero” por “Eósforo” pode alterar o valor cognitivo da frase. O diagnóstico de Frege é que o valor cognitivo depende também do significado das expressões envolvidas. É porque “Héspero” tem um significado diferente de “Eósforo” que um sujeito pode coerentemente acreditar em (16) e rejeitar (17).

Infelizmente, Frege não disse muito sobre o que exatamente é o significado, sentido ou modo de apresentação de uma expressão. Dois pontos devem ser ressaltados. Primeiro, de acordo com Frege é o modo de apresentação que determina o referente de uma expressão, quando esta tem um referente. Assim, “Héspero” e “Eósforo” têm o mesmo referente porque seu significado de algum modo determina o mesmo objeto como referente de ambas. Portanto, a teoria fregeana é uma versão de teoria da referência indireta. Segundo, ele parece ter aceitado que expressões podem ter sentido mesmo quando falham em se referir a algo.¹¹⁶ É isto que ocorre com nomes vazios, como “Sherlock Holmes” por exemplo. Aliás, esta é outra suposta vantagem da teoria fregeana sobre o referencialismo, dado que o último implicaria a conclusão aparentemente absurda de que nomes vazios são destituídos de significado.

Tudo isto levanta um conjunto de perguntas. Qual exatamente é o significado de um nome? Até agora distinguimos entre significado e referente, mas não dissemos qual é o significado de um nome. Esta é uma instância do problema descritivo do significado. Como exatamente o significado de um nome

¹¹⁵Estes contextos – em que a substituição de uma expressão por outra com o mesmo referente não altera o valor de verdade da frase completa – são chamados de “contextos referencialmente transparentes”. Há uma classe especial de contextos, chamados “contextos referencialmente opacos” em que a substituição de uma expressão por outra com o mesmo referente pode sim alterar o valor de verdade da frase completa. Não levo em conta este último tipo de contexto aqui. Para uma apresentação introdutória e muito clara da crítica de Frege ao referencialismo sobre os nomes próprios, que leva em conta os contextos referencialmente opacos, veja Juliana Faccio Lima (2016). Para uma discussão crítica dos casos de opacidade referencial, veja João Branquinho (2003).

¹¹⁶Digo “parece” porque isto é por vezes colocado em causa. Evans (1982), por exemplo, rejeita que Frege tenha sequer pretendido acomodar a possibilidade de sentido sem referente. Em meu Salles (2019), aponto pelo menos uma dificuldade para a interpretação de Evans.

determina seu referente? Ainda precisamos dizer algo sobre o mecanismo envolvido na determinação do referente. Esta é uma instância do problema fundacional da referência.

12. Descritivismo clássico e nomes próprios

O descritivismo clássico fornece uma resposta às perguntas com as quais encerramos a seção anterior. Esta teoria é usualmente atribuída a Frege (1892) e Russell (1912, p. 29; 1905; 1972, p. 29), mas não é claro que qualquer um deles tenha defendido exatamente isto (COSTA, 2009). De acordo com o descritivismo clássico, o significado de um nome é o significado de uma descrição definida associada pelo falante ao mesmo, e é o significado da descrição que determina o referente do nome. Neste contexto, o primeiro passo é entender qual o significado das descrições definidas; ou seja, entender como descrições contribuem para o significado das frases completas que as contém. Isto geralmente é feito por meio da *teoria das descrições definidas* desenvolvida por Bertrand Russell (1905). De acordo com esta teoria, devemos interpretar uma frase da forma “o F é G” como uma frase existencial complexa do seguinte tipo: “*existe um e apenas um F*, e quem quer que seja F é G”. Uma frase como (18), por exemplo, deve ser interpretada em termos de (19).

18. O fundador da lógica formal é sábio.

19. *Existe um e apenas um fundador da lógica formal*, e quem quer que tenha fundado a lógica formal é sábio.

A parte em itálico representa a contribuição feita pela descrição para o significado da frase completa. Grosso modo, descrições definidas são usadas para afirmar a existência de um e apenas um indivíduo que satisfaz certa propriedade ou conjunto de propriedades. Fazendo uso dessa teoria, alcançamos uma interessante explicação da referência das descrições. O referente – ou *denotação*, como é mais comum no caso das descrições – de uma descrição definida “o F” será quem quer que seja o único objeto a possuir a propriedade F. A denotação de “o fundador da lógica formal”, por exemplo, é quem quer que seja o único indivíduo (se há algum) a ter fundado a lógica

formal.¹¹⁷ Com isto podemos finalmente explicar a referência dos nomes. O referente de um nome será justamente a denotação da descrição associada a ele. Se a descrição que associa ao nome “Aristóteles” é “o fundador da lógica formal”, então o referente deste nome, como usado por mim, será o único indivíduo a ter fundado a lógica formal. Se não existir alguém que fundou a lógica formal, ou existir mais de uma pessoa que o tenha feito, então falho em me referir a algo pelo nome. De um modo mais geral, o referente de um nome N será sempre a denotação de uma descrição associada “o F”. E o referente de uma descrição “o F” é o único objeto (se existir algum) a possuir a propriedade F.

O descritivismo clássico é elegante por muitas razões. Para começo de conversa, ele nos permite explicar as diferenças entre (16) e (17) acima. Como esperado, o diagnóstico seria que as diferenças se devem ao fato de os nomes relevantes serem associados a descrições diferentes e, portanto, terem significados diferentes. Por exemplo, “Héspero” pode ser associado à descrição “a estrela vespertina”, enquanto “Eósforo” é associado à descrição “a estrela matutina”. Uma vez que estas descrições têm significados distintos, os nomes “Héspero” e “Eósforo” também têm significados distintos. Conseqüentemente, (16) e (17) – que seriam respectivamente equivalentes a “a estrela vespertina é a estrela vespertina” e “a estrela vespertina é a estrela matutina” – exprimem proposições diferentes. Isso se acomoda ao fato de que apenas (16) é uma identidade trivial, e de que (16) e (17) têm valores cognitivos diferentes. Além disto, o descritivismo clássico parece congruente com nossas práticas de adquirir e de ensinar nomes. No mínimo, uma boa forma de ensinar alguém a usar um nome é introduzi-lo juntamente com uma descrição definida. Do mesmo modo, no mínimo, uma boa forma de aprender a usar um nome é recebê-lo juntamente com uma descrição definida. Por exemplo, quando digo a você que Carmensiva é a minha mãe, lhe ensino a usar o nome “Carmensiva” para se referir a minha mãe, e você aprende a usá-lo para se referir a ela.

Todavia, mesmo que aceitemos a teoria das descrições de Russell, não é de todo claro que o significado e a referência dos nomes possam ser explicados em termos dela. No que segue, considero brevemente dois argumentos importantes contra o descritivismo clássico. O argumento modal, de

¹¹⁷Uma breve introdução à teoria das descrições de Russell é meu Salles (2010). Para uma introdução muito mais completa veja Neale (1990).

Saul Kripke, coloca em causa a solução para o problema descritivo do significado dos nomes; enquanto o argumento semântico, de Kripke e Keith Donnellan, coloca em causa a solução para o problema fundacional da referência.¹¹⁸

13. Argumento modal e designadores rígidos

De acordo com descritivistas clássicos, o significado de um nome próprio é fornecido pela descrição definida associada ao mesmo, de modo que nomes são equivalentes a descrições. O argumento modal de Kripke (1977/80, p. 48-49) pretende mostrar que nomes não são equivalentes a descrições, porque nomes e descrições se comportam de modo diferente em contextos modais (ou seja, contextos envolvendo possibilidade e necessidade).

Comece pensando na seguinte frase:

20. Se Aristóteles existe, então Aristóteles é Aristóteles.

(20) é necessariamente verdadeira. Não é possível que Aristóteles exista e Aristóteles não seja Aristóteles. Claro que Aristóteles poderia existir e ter outro nome, mas o que (20) afirma é que se ele existe, então ele é ele mesmo. Isso não poderia ser falso.

Agora retire a última ocorrência do nome “Aristóteles” de (20), e a substitua pela descrição “o fundador da lógica formal”. O resultado é (21).

21. Se Aristóteles existe, então Aristóteles é *o fundador da lógica formal*.

Repare que, diferentemente de (20), (21) não é necessariamente verdadeira. Aristóteles poderia ter existido e, ainda assim, não ter fundado a lógica formal. Ele poderia, por exemplo, ter morrido logo após nascer, ou ter se

¹¹⁸Atualmente, existem três argumentos principais contra o descritivismo: o *argumento modal*, o *argumento epistemológico* e o *argumento semântico*. Tanto o argumento modal como o epistemológico colocam em causa a solução para o problema descritivo do significado, e por isto decidi tratar apenas um deles aqui. A escolha pelo argumento modal se deve ao fato de parecer-me bem mais influente do que o epistemológico.

dedicado exclusivamente à carpintaria, sem gastar seu tempo com coisas tediosas como a lógica. Portanto, (20) é necessariamente verdadeira, mas (21) não. Se isso é assim, então (20) e (21) devem expressar proposições diferentes. Mas a única diferença entre ambas é que nós substituímos uma das ocorrências do nome “Aristóteles” (em (20)) pela descrição “o fundador da lógica formal” (em (21)). Deste modo, a única explicação para a diferença de significado entre ambas é que o nome e a descrição contribuem de um modo diferente para o significado das frases. Por outras palavras, o nome “Aristóteles” tem um significado diferente da descrição “o fundador da lógica formal”.

O problema é que o argumento pode ser generalizado. Tome as descrições normalmente associadas ao nome “Aristóteles”: “o fundador da lógica formal”, “o autor da *Metafísica*”, “o discípulo mais importante de Platão” etc. Se colocarmos qualquer uma delas no lugar da última ocorrência de “Aristóteles” em (20), o resultado será uma proposição que, diferentemente de (20), não é necessariamente verdadeira. Assim, pelo mesmo argumento, nenhuma dessas descrições é equivalente ao nome “Aristóteles”. Em conclusão, a solução descritivista para o problema descritivo do significado dos nomes próprios deve ser rejeitada.

Kripke vai além e fornece um argumento adicional. Este segundo argumento apela ao fato de que nomes, mas não descrições, são designadores rígidos. O conceito de *designador rígido* é geralmente definido em termos do conceito de *mundos possíveis*. O idioma dos mundos possíveis é usado para tornar mais clara a nossa conversa sobre necessidades e possibilidades. Grosso modo, um mundo possível é um modo como as coisas podem ser. O modo como as coisas efetivamente são é também um modo como elas podem ser e, assim, é também um mundo possível; ao qual chamamos “mundo atual”. Repare que o mundo atual (o modo como as coisas efetivamente são) não inclui apenas o presente. De fato, por *mundo atual* entende-se todo o nosso universo, incluindo seu passado, presente e futuro com tudo que há nele. Agora, podemos traduzir a nossa conversa sobre possibilidades e necessidades para uma linguagem que use a noção de mundos possíveis. Dizer que algo é possível é dizer que algo é o caso em pelo menos um mundo possível. Dizer que algo é necessário é dizer que é o caso em todos os mundos possíveis.

Vimos que (21) não é necessariamente verdadeira. Isto significa que (21) não é verdadeira em todos os mundos possíveis. Por exemplo, (21) é falsa em

mundos nos quais Aristóteles existe, mas não é o fundador da lógica formal. Uma vez que mesmo nestes mundos Aristóteles continua sendo Aristóteles, (20) continua sendo verdadeira. A proposição expressada por (20) é verdadeira em todos os mundos possíveis, é necessariamente verdadeira. Enquanto o valor de verdade de (21) varia de mundo para mundo, o valor de verdade de (20) não varia. Kripke pensa que algo análogo acontece com descrições e nomes próprios. Enquanto o referente de nomes normalmente não varia de mundo para mundo, o de descrições normalmente varia.¹¹⁹ Começemos pelo caso dos nomes.

Para a frase “Aristóteles é sábio” ser verdadeira em um mundo possível *m*, basta que a pessoa de fato referida pelo nome “Aristóteles” seja sábia em *m*. Não há qualquer problema adiante sobre quem é o referente de “Aristóteles” em *m*. Este nome se referirá em *m* à mesma pessoa a quem ele se refere no mundo atual. Por outras palavras, o referente de “Aristóteles” é constante de mundo para mundo. Por isto, para avaliarmos a verdade da proposição expressa por esta frase com respeito a uma situação possível, não é preciso procurar pelo referente do nome naquela situação.

Descrições funcionam de um modo diferente. A frase “O vocalista do Radiohead é um homem” é verdadeira. Mas consideremos um mundo possível onde Marília Mendonça é a vocalista do Radiohead. Neste mundo, a frase acima é falsa. Se quisermos avaliar a verdade desta frase com respeito a uma dada situação possível, precisaremos descobrir quem é o referente da descrição naquela situação. De modo geral, quando avaliamos a verdade de uma frase da forma “o F é G” com respeito a uma situação possível, temos de procurar pelo referente de “o F” naquela situação. Isto ocorre porque o referente de uma descrição pode mudar de um mundo possível para outro; portanto não é constante. Para expressar esta diferença, diz-se que nomes, mas não descrições, são *designadores rígidos*. A noção intuitiva de designador rígido é a seguinte: um termo *t* é um designador rígido se, e somente se, designa o mesmo indivíduo em todos os mundos possíveis onde ele existe, e designa nada mais em qualquer

¹¹⁹A expressão “normalmente” não é usada ao acaso. O referente de algumas descrições definidas não varia de mundo para mundo. Por exemplo, o referente ou denotação da descrição “o indivíduo que atualmente é o fundador da lógica formal” é constante. Este tipo de descrição, chamado de “descrição rigidificada”, pertence à classe do que mais à frente chamaremos de “designadores rígidos”.

outro mundo possível.¹²⁰

Isto explica *parcialmente* por que (21) é falsa em alguns mundos possíveis. Enquanto o referente de “Aristóteles” é constante, o referente de “o fundador da lógica formal” não é. Assim, embora Aristóteles seja o referente desta descrição no mundo atual, existem mundos nos quais Aristóteles existe, mas a descrição “o fundador da lógica formal” não o denota. Ora, quando usarmos o nome “Aristóteles” para falar de um desses mundos, nos referiremos ao próprio Aristóteles, já que o referente desse nome não varia de mundo para mundo. Mas quando usarmos a descrição associada, não denotaremos Aristóteles. O resultado é que (21) será falsa neste mundo.

Em resumo, a versão mais sofisticada do argumento de Kripke pode ser exposta como segue: nenhum termo que seja um designador rígido pode ser equivalente a (ter o mesmo significado de) um termo que não seja um designador rígido. Assim, nenhum nome que seja um designador rígido pode ser equivalente a uma descrição não rígida. Dado que Kripke está confiante de que nomes geralmente são rígidos e as descrições mais comuns associadas a eles geralmente não são rígidas, fica difícil aceitar o descritivismo.¹²¹

14. Argumento semântico e teoria causal

O próximo argumento que consideramos coloca em xeque a resposta descritivista para o problema fundacional da referência.

Deixe “o fundador da lógica formal” ser a descrição associada por João ao nome “Aristóteles”. Neste caso, há apenas duas opções: ou existe um único indivíduo que fundou a lógica formal – e este será o referente de “Aristóteles” na boca de João – ou João falha em se referir a qualquer objeto por meio deste nome. Isto significa que, se João de fato se refere a algo por meio do nome, então ele não pode estar errado sobre o fato de que seu referente é o fundador

¹²⁰Existem tipos diferentes de designadores rígidos (BROCK, 2004, p. 283), mas não vou entrar nestas distinções aqui.

¹²¹ O argumento modal talvez seja o mais discutido nos últimos anos, e foram desenvolvidas pelo menos duas versões novas do descritivismo para lidar com este argumento: o *descriptivismo de escopo amplo* (SOSA, 2001) e o *descriptivismo rigidificado* (NELSON, 2002, CID, 2010). Estas versões de descritivismo são criticadas em Soames (1998) e Branquinho (2001).

da lógica formal. Por isto, podemos dizer que João *sabe* que o referente de “Aristóteles” é o fundador da lógica formal. Isto indica que a referência por nomes envolve um forte requisito epistêmico. O requisito pode ser colocado como segue: um falante S tem sucesso em se referir a um objeto *x* através de um nome N somente se *sabe* que o referente de N é o único a possuir uma determinada propriedade. O argumento semântico coloca em causa este requisito epistêmico.

Não é muito difícil pensar em exemplos de falantes que (i) são referencialmente bem-sucedidos através de um nome próprio, mas (ii) não sabem que o referente do nome é o único a possuir certa propriedade. Suponha que Maria nunca tenha ouvido falar de Aristóteles, e que um professor lhe conte que Aristóteles é o mestre de Sócrates. Uma vez que a única descrição que Maria associa ao nome é falsa de Aristóteles, ela não tem o tipo de conhecimento exigido por descritivistas. Ainda assim, parece-nos que Maria é capaz de se referir a Aristóteles por meio de “Aristóteles”. Quando ela diz “Aristóteles é o mestre de Sócrates”, está a dizer uma falsidade *sobre Aristóteles* e, portanto, é bem-sucedida em seu ato de referência. Exemplos deste tipo podem ser facilmente multiplicados (KRIPKE, 1972/80, p. 95; DONNELLAN, 1972). Resultado: o sucesso da referência por um nome não depende do tipo de conhecimento exigido pelos descritivistas. A solução descritivista do problema fundacional da referência dos nomes não funciona.¹²²

Exemplos como este indicam que uma solução para o problema fundacional da referência dos nomes deveria acomodar o fato de que falantes significativamente ignorantes sobre o objeto nomeado podem, ainda assim, ser referencialmente bem-sucedidos através do nome. Kripke, um dos autores do argumento semântico, nos apresentou um esboço de tal solução. De acordo com este esboço, a existência de um elo causal apropriado entre o usuário do nome e o objeto referido é uma condição suficiente para o sucesso da referência. A ideia é bem simples. Há muitos séculos, um menino foi batizado de “Aristóteles” por pessoas que estavam em contato com ele. Estas pessoas então difundiram o nome para outras, que difundiram para outras, e assim sucessivamente até nos alcançar. A sugestão de Kripke é que este processo de difusão do nome deve ser interpretado como uma cadeia causal que remete àquele indivíduo batizado:

¹²²Respostas descritivistas ao argumento semântico podem ser encontradas em Kroon (1983), Loar (1976) e Costa (2011, 2013).

o próprio Aristóteles. Quando uso o nome “Aristóteles” hoje, me refiro àquele indivíduo precisamente porque sou um membro desta cadeia. Não é necessário, portanto, qualquer conhecimento de que o referente do nome é o único a possuir esta ou aquela propriedade. Obviamente, o mesmo se aplica à Maria de nosso exemplo acima. Este tipo de explicação da referência é conhecido como *teoria causal da referência*, e foi posteriormente desenvolvido por filósofos como Devitt (1981. DEVITT; STERELNY, 1999) e Andrea Bianchi (2015). Atualmente, é considerado por muitos a nova ortodoxia quanto ao problema fundacional da referência.¹²³

Por fim, repare que a teoria causal é uma resposta ao problema fundacional da referência, e nada nos diz sobre o significado dos nomes. A despeito disso, ela é convenientemente compatível com a resposta referencialista ao problema descritivo do significado, e pode completar a última. Afinal, para além da implicação negativa de que o referente de um nome não é determinado por seu significado, o referencialismo nada diz sobre como o referente é determinado. A teoria causal pode fornecer a parte que faltava.

Vimos que a tese referencialista dos nomes próprios tem dificuldades em explicar algumas diferenças relevantes entre “Héspero é Héspero” e “Héspero é Eósforo”, e que Frege acreditava que a melhor forma de as explicar era postulando uma distinção entre significado e referente. Frege pensava que o significado determinava o referente, mas não explicou bem como isto ocorria. O descritivismo clássico fornece uma resposta aos problemas deixados em aberto, mas também enfrenta problemas difíceis. O argumento modal coloca em xeque a sua solução para o problema descritivo do significado, enquanto o argumento semântico coloca em xeque sua solução para o problema fundacional da referência. A teoria causal aparece como uma solução alternativa para o último problema, mas deixa em aberto a solução do problema descritivo do significado dos nomes. A despeito disto, ela é compatível com o referencialismo, e pode ser um complemento bem-vindo. Ainda não alcançamos qualquer coisa próxima de

¹²³Como esperado, a teoria causal também encontra seus objetores. O artigo clássico aqui é Evans (1973). Objeções interessantes à versão mais recente de Bianchi podem ser encontradas em Jessica Pepp (2018). Peter Unger (1983) é uma fonte instigante de objeções à tentativa de aplicar a teoria causal ao problema fundacional da referência dos termos gerais.

um consenso sobre qual o melhor caminho, e referencialismo, descritivismo e teoria causal aparecem no debate contemporâneo em versões sofisticadas como opções genuínas de solução para os problemas relevantes. Acessar os detalhes deste debate está além do objetivo deste texto, e me dou por satisfeito se o que foi apresentado aqui servir como uma amostra da discussão filosófica envolvendo os conceitos de *significado* e *referência*.

15. Paradoxo sorites e a incoerência da linguagem natural

Na introdução, afirmei que tudo corre relativamente bem com nossa linguagem natural. Isto é bastante intuitivo, e filósofos da linguagem frequentemente trabalham sob esta suposição. Infelizmente, contudo, há argumentos que a colocam em causa. Um destes argumentos, descoberto no século IV a.c. pelo mestre dos paradoxos, Ebulides de Mileto, é o *paradoxo sorites*. Um paradoxo é um argumento aparentemente válido com premissas aparentemente verdadeiras e conclusão aparentemente absurda ou contraditória. O paradoxo sorites coloca em xeque a coerência da maior parte das expressões da linguagem natural, colocando em causa o otimismo com o qual comecei este texto.

Comece considerando a expressão “careca”. Uma pessoa com 0 fios de cabelo em seu couro cabeludo é careca. Ora, se uma pessoa com 0 fios é careca, então uma pessoa com 1 fio também é careca. Afinal, seria absurdo supor que a adição de um único fio de cabelo pode tornar um careca em um não careca. Portanto, uma pessoa com 1 fio de cabelo em seu couro cabeludo é careca. Se uma pessoa com 1 fio é careca, então uma pessoa com 2 fios também o é. Afinal, a adição de um único fio de cabelo não pode tornar um careca em um não careca. Portanto, uma pessoa com 2 fios é careca. Se uma pessoa com 2 fios é careca, então uma pessoa com 3 fios também o é. Portanto, uma pessoa com 3 fios é careca. Continue passo a passo este procedimento, e poderá mostrar que qualquer pessoa é careca, mesmo um indivíduo com 10.000 fios de cabelo na cabeça.

Agora façamos o procedimento inverso. Sabemos que uma pessoa com 10.000 fios de cabelo em seu couro cabeludo é não careca. Mas se uma pessoa com 10.000 fios é não careca, então uma pessoa com 9.999 fios também é não careca. Afinal, a remoção de um único fio de cabelo não pode tornar uma pessoa

não careca em um careca. Portanto, uma pessoa com 9.999 fios de cabelo é não careca. Ora, se uma pessoa com 9.999 fios é não careca, então uma pessoa com 9.998 também é não careca. Portanto, uma pessoa com 9.998 fios é não careca. Seguindo este procedimento, poderá mostrar que qualquer indivíduo é não careca, mesmo alguém com 0 fios de cabelo na cabeça.

Cada argumento acima parte de uma suposição inocente e alcança, por meios aparentemente legítimos, uma conclusão absurda. O primeiro parte da suposição de que uma pessoa com 0 fios é careca, e conclui que uma pessoa com 10.000 fios também o é. O segundo parte da suposição de que uma pessoa com 10.000 fios é não careca, e conclui que uma pessoa com 0 fios também é não careca. Estes argumentos são instâncias do *paradoxo sorites*. Conjuntamente, os dois argumentos acima nos permitem mostrar, para cada indivíduo, que ele é simultaneamente careca e não careca. O termo geral “careca” é incoerente.

Se o problema dissesse respeito especificamente à expressão “careca”, poderíamos retirá-la de nosso vocabulário e seguir em frente com nossas vidas. Infelizmente, contudo, o mesmo problema surge para a maior parte dos termos gerais de nossa linguagem. A título de um exemplo extra, considere a expressão “barato”, aplicada a pizzas gostosas. Uma pizza gostosa de 0 centavos é barata. Se uma de 0 centavos é barata, uma de 1 centavo também o é. (Continue o procedimento). Logo, uma pizza de R\$ 70.000.000.000 é barata. Agora caminhemos na direção inversa. Uma pizza gostosa de R\$ 70.000.000.000 não é barata. Se uma de R\$ 70.000.000.000 não é barata, então uma de R\$ 69.999.999.999,99 também não é barata. (Continue o procedimento). Logo, uma pizza gostosa de 0 centavos também não é barata. Com o paradoxo sorites podemos mostrar, para qualquer pizza gostosa, que ela é barata e não barata. Expressões como “alto”, “baixo”, “grande”, “médio”, “pequeno”, “grito”, “caro”, “barato”, “forte”, “fraco”, “rico”, “pobre”, etc. são todas suscetíveis ao paradoxo sorites, e portanto todas incoerentes no sentido forte de que tanto elas como suas respectivas negações se aplicam a cada objeto dos domínios relevantes. Isto significa que a maior parte dos termos gerais de nossa linguagem são incoerentes, sem contar que o paradoxo também se aplica a termos singulares e quantificadores. Em suma, o sorites coloca em xeque a coerência de nossa linguagem como um todo.

O paradoxo sorites admite muitas versões diferentes, e mesmo a versão

apresentada acima, conhecida como *versão condicional*, pode ser formulada de forma muito mais precisa. Para uma apresentação introdutória com foco nas três versões principais do sorites, incluindo a apresentada acima, veja meu Salles (2016); Dominic Hyde (2011) apresenta ainda outras versões interessantes. Seja qual for a versão em jogo, há três formas de solucionar um paradoxo, incluindo o sorites. A primeira é simplesmente aceitar a conclusão. A segunda é apontar algum erro na derivação da conclusão, sustentando que o argumento não é realmente válido. A terceira é rejeitar a verdade de alguma das premissas do argumento. Poucos filósofos estão dispostos a adotar a primeira saída, alguns adotam a segunda e muitos (eu incluso) adotam a terceira. Ao longo das últimas décadas, muitas soluções diferentes foram propostas, geralmente associadas às teorias da vagueza. Usualmente, considera-se que o fenômeno por trás do paradoxo sorites é a vagueza, de modo que compreendê-la se tornou central para resolver o paradoxo. Infelizmente, ainda não chegamos a algo sequer perto de um consenso sobre qual a teoria correta da vagueza. Uma exposição curta e bem introdutória do problema e de algumas soluções é Sainsbury (2009). Discussões mais detalhadas e também muito competentes são Timothy Williamson (1994) e Rosanna Keefe (2000).

16. Considerações finais

Tivemos a oportunidade de considerar alguns dos principais problemas da filosofia da linguagem, introduzir parte de seus conceitos fundamentais e investigar de forma um pouco mais detalhada o exemplo específico dos nomes próprios. Com isto, pode-se ter uma ideia do tipo de problemas filosóficos relacionados à linguagem, em particular aos fenômenos da referência e do significado, e de como lidamos com eles em filosofia. Todavia, é importante ressaltar que há uma miríade de problemas que foram ignorados nesta introdução. Em especial, nada foi dito sobre as teorias fundacionais do significado, que têm ocupado posição central nos debates contemporâneos. O conceito de *atos de fala* também foi ignorado, assim como a distinção entre *internalismo* e *externalismo* semântico, e os interessantes paradoxos semânticos.¹²⁴ De fato, eu poderia gastar algumas páginas aqui apenas

¹²⁴ O leitor poderá encontrar mais sobre o problema fundacional do significado

mencionando o que não foi incluído em nossa discussão. Mas basta que o leitor esteja ciente de que viu apenas a ponta do iceberg, ou talvez a ponta da ponta. Isto basta para um começo.

Questões para revisão

1. O que é uma teoria da referência, e o que distingue os problemas descritivo e fundacional da referência?
2. Aponte uma diferença entre termos gerais e termos singulares.
3. O que é uma teoria do significado, e o que distingue os problemas descritivo e fundacional do significado?
4. Explique, com suas próprias palavras, o princípio da composicionalidade.
5. Conteúdo e significado linguístico são a mesma coisa? Justifique sua resposta. (Sugestão: tome como exemplo o indexical “hoje” e explique a diferença entre seu significado linguístico e seu conteúdo).
6. Explique, com suas próprias palavras, a distinção entre o que é dito e o que é implicitado. Forneça exemplos originais a favor da distinção.
7. Explique brevemente o referencialismo e o argumento de Frege contra o mesmo.
8. Explique brevemente (a) o descritivismo clássico, (b) o argumento modal e (c) o argumento semântico.
9. A palavra “sorites” vem do grego “soros”, que significa *monte* ou *pilha*. Isto se deve ao fato de que ele foi inicialmente formulado para expressão “monte”. Consegue formular uma versão do paradoxo para esta expressão?

Questões para discussão

1. Em sua opinião, qual é a importância de compreendermos os fenômenos do significado e referência?
2. Acha que a condição *deno máximo um* é suficiente para a distinção entre

praticamente em qualquer livro de introdução à filosofia da linguagem, em especial o já indicado Lycan (2000). Sobre atos de fala, uma breve e boa introdução é Bach (2006a). Apesar de não ser uma introdução, Searle (1969) é sempre claro, além de já ser um dos clássicos sobre o tema. Sobre os paradoxos semânticos veja Santos (2014).

- termos gerais e singulares? Justifique a sua resposta.
3. É capaz de apontar algum outro sentido de *significado*, que seja distinto dos dois sentidos anteriores (conteúdo e significado linguístico)?
 4. Construa exemplos de situações cotidianas nas quais o que é dito não coincide com o que é implicado, e tente especular sobre como ocorre o processo de inferência do que é implicado em cada caso.
 5. Acha que o descritivismo clássico poderia ser aplicado na explicação da referência e significado de outros termos singulares, como indexicais puros e verdadeiros demonstrativos? Justifique a sua resposta.
 6. Nomes próprios, mas não as descrições definidas mais comumente associadas a eles, são designadores rígidos. Consegue pensar em outras categorias de expressões que são designadores rígidos?
 7. Especule sobre quais seriam as consequências caso maior parte de nossas expressões fossem realmente incoerentes, no sentido apontado pelo paradoxo sorites.

Sugestão de leitura

Uma introdução à filosofia da linguagem não apenas competente, mas também muito divertida, é Lycan (2000). Devitt e Sterelny (1999) também pode ser de grande ajuda; embora Kemp (2013) seja mais atual. Sofia Miguens (2007) faz uma introdução ao tema com foco na discussão que se estende do final do século XIX até meados do século XX, e fornece uma boa discussão de parte importante da história da filosofia da linguagem. Imaguire e Schirn (2008) abordam de forma relativamente introdutória vários temas centrais da filosofia da linguagem contemporânea. Três fontes online interessantes são a *Stanford Encyclopedia of Philosophy*, o *Compêndio em Linha de Problemas de Filosofia Analítica* e o site criticanarede.com; onde o leitor encontrará amplo acervo de material enciclopédico e introdutório sobre as mais variadas áreas da filosofia, incluindo a filosofia da linguagem. De resto, cada seção do presente texto já indicou fontes adicionais, introdutórias e não introdutórias, sobre os temas discutidos.

Referências

- BACH, Kent. (1994). Conversational Implicature. *Mind & Language*, 9 No. 2, pp. 124-162.
- BACH, Kent. (1997). The semantics-pragmatics distinction: What it is and why it matters. *Linguistische Berichte* 8: 33–50.
- BACH, Kent (2006a). Speech Acts and Pragmatics. In: DEVITT, M.; HANLEY, R. (ed.). *The Blackwell Guide to the philosophy of language*. Malden: Blackwell.
- BACH, Kent. (2006b). The Excluded Middle: Semantic Minimalism without Minimal Propositions. *Philosophy and Phenomenological Research*, LXXIII, No. 2, pp. 436-442.
- BIANCHI, Andrea (2015). Reference and Repetition. In: BIANCHI, Andrea (org.). *On Reference*. Oxford: Oxford University Press, p. 93-107.
- BRANQUINHO, João (2001). Sobre o Descritivismo Atualizado aos Argumentos Modais?
- BRANQUINHO, João. (2003). Opacidade Referencial: Afinal Não Há Assim Tantos Casos! In *Razões e Emoção. Miscelânea de Estudos em Homenagem a Maria Helena Mira Mateus*. Lisboa: Imprensa Nacional - Casa da Moeda.
- BROCK, Stuart (2004). The Ubiquitous Problem of Empty Names. *The Journal of Philosophy*, v. 101, n. 6, p. 277-298.
- CAPPELEN, Herman; LEPORE, Ernie (2005). *Insensitive Semantics*. Oxford: Blackwell.
- CARSTON, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.
- CID, Rodrigo. (2010). “Descritivismo atualizado e protegido: uma resposta a Célia Teixeira”. *Aproximação*: v. 3, p. 112-119.
- COSTA, Cláudio F. (2009). Teorias Descritivistas dos Nomes Próprios. *Dissertatio*, v. 30, p. 185-195.
- COSTA, Cláudio F. (2011). A Meta-Descriptivist Theory of Proper Names. *Ratio*, v. 24, n. 3, p. 259-281.
- COSTA, Cláudio F. (2013). Como os nomes próprios realmente referem. *Investigação Filosófica*, v. 4, n. 1, p. 1-27.
- DEVITT, Michael (1981). *Designation*. Columbia University Press.
- DEVITT, Michael; STERELNY, Kim (1999). *Language and Reality: an introduction*

- to the philosophy of language. Oxford: Blackwell.
- EVANS, Gareth (1973). The causal Theory of Names. In: DAVIDSON, Matthew (org.). *On Sense and Direct Reference*. The McGraw-Hill Companies, 2007, p. 528-543.
- EVANS, Gareth (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- FREGE, Gottlob (1892). Sobre o Sentido e a referência. In: *Lógica e Filosofia da Linguagem*. São Paulo: Cultrix, 1978. Tradução de Paulo Alcoforado.
- GRICE, H. Paul (1967). *Lógica e Conversação*. *Crítica*, 2016. Tradução Matheus Silva.
- HAACK, Susan (1978). *Filosofia das Lógicas*. São Paulo: Unesp, 2002. Tradução: Cezar Augusto Mortari & Luiz Henrique de Araújo.
- HYDE, Dominic (2011). The sorites paradox. In: RONZITTI, G. *Vagueness: A guide* (pp. 1-17). Springer Science & Business Media.
- IMAGUIRE, Guido (2007). A Negação e as Árvores Conceituais. *O que nos Faz Pensar*, n. 23.
- IMAGUIRE, Guido; SCHIRN, Matthias (2008). *Estudo em Filosofia da Linguagem*. São Paulo: Edições Loyola.
- KAPLAN, David (1977). Demonstratives. In: ALMOG, Joseph; PERRY, John; Howard WETTSTEIN, Howard (orgs.). *Themes from Kaplan*. Oxford: Oxford University Press, 1989, p. 481-563.
- KAPLAN, David (1989). Afterthoughts. In: ALMOG, Joseph; PERRY, John; WETTSTEIN, Howard (orgs.). *Themes from Kaplan*. Oxford: Oxford University Press, 1989, p. 565-614.
- KEMP, Gary (2013). *What is this Thing Called Philosophy of Language?* London: Routledge.
- KRIPKE, Saul A. (1972/1980). *Naming and Necessity*. Oxford: Basil Blackwell, 1990.
- KROON, Frederick W. (1983). The Problem of 'Jonah': how not to argue for the causal theory of reference. *Philosophical Studies*, v. 43, n. 2, p. 281-299.
- MEINONG, Alexius (1904). The theory of objects. In: CHISHOLM, Roderick M. (org.) *Realism & the background of phenomenology*. Free Press, 1960, p.76-117.
- MIGUENS, Sofia (2007). *Filosofia da Linguagem*. Faculdade de Letras da Universidade do Porto.

- NEALE, Stephen (1990). *Descriptions*. The Massachusetts Institute of Technology Press (MIT).
- NELSON, Michael (2002). Descriptivism Defended. *Noûs*, v. 36, n. 3, p. 408-435.
- LIMA, Juliana F. (2016). O Enigma de Frege. *Fundamento*, n. 12.
- LOAR, Brian (1976). The Semantics of Singular terms. In: DAVIDSON, Matthew (org.). *On sense and direct reference*. The McGraw-Hill Companies, 2007, p. 496-516.
- LYCAN, William G. (2000). *Philosophy of Language: a Contemporary Introduction*. New York: Routledge.
- PEPP, Jessica (2018). What Determines the Reference of Names? What Determines the Objects of Thought. *Erkenntnis*, v. 84, n. 4, p. 741-759.
- RECANATI, F. (2001). What is said. *Synthese* 128: 75–91
- RECANATI, F. (2004). *Literal meaning*. Cambridge: Cambridge University Press.
- RUSSELL, Bertrand (1905). On Denoting. *Mind*, v. 14, n. 56, p. 479-493.
- RUSSELL, Bertrand (1912). *The Problems of Philosophy*. Oxford: Oxford University Press, 2001
- RUSSELL, Bertrand (1972). *The Philosophy of Logical Atomism*. Routledge.
- SAINSBURY, Mark. (2006). The essence of Reference. In: LEPORE, Ernest. and SMITH, Barry C. (orgs.). *Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press, p. 393-422.
- SAINSBURY, Sainsbury, Mark (2009). *Paradoxes*. Cambridge University Press.
- SALLES, Sagid (2010). Introdução à Teoria das Descrições de Russell. *Criticana rede*. Disponível em: <https://criticanarede.com/descricoes.html>
- SALLES, Sagid (2015). Fazendo de Conta que Vulcano não Existe. *Philosophos*, v. 20, n. 2, pp. 171-196.
- SALLES, Sagid (2016). O Problema da Vagueza. *Fundamento*, No.12, pp. 139-174.
- SALLES, Sagid (2019). Nota sobre o Frege de Evans. *Investigação Filosófica*, v. 10, n. 2, p. 39-46.
- SANTOS, Ricardo. (2014). Paradoxos Semânticos. In: BRANQUINHO, J.; SANTOS, R. (Eds.), *Compêndio em Linha de Problemas de Filosofia Analítica* (pp. 1-41). Lisboa.
- SEARLE, John R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.
- SEARLE, John R. (1979). O Estatuto Lógico do Discurso Ficcional. In: *Expressão e Significado*. Martins Fontes, 1995. Tradução de Ana Cecília G. A. de

- Camargo e Ana Luiza Marcondes Garcia.
- SOAMES, Scott (1998). The Modal Argument: Wide Scope and Rigidified Descriptions. *Noûs*, v. 32, n. 1, p. 1–22.
- SOSA, David (2001). Rigidity in the Scope of Russell's Theory. *Noûs*, v. 35, n. 1, p. 1-38.
- STALNAKER, Robert (1997). Reference and Necessity. In: HALE, Bob; WRIGHT, Crispin (orgs.). *A Companion to the Philosophy of Language*. Oxford: Blackwell, 1998, p. 534-554.
- STRAWSON, Peter F. (1952). *Introduction to Logical Theory*. London: Methuen & Co Ltd.
- STRAWSON, Peter F. (1959). *Individuals*. London: Routledge.
- SZABÓ, Zoltán Gendler (2020). Compositionality. In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). Forthcoming URL = <https://plato.stanford.edu/archives/fall2020/entries/compositionality/>
- TRAVIS, Charles (1997). Pragmatics. In: HALE, Bob; WRIGHT, Crispin (orgs.). *A Companion to the Philosophy of Language*. Oxford: Blackwell, 1998, p. 87-107).
- UNGER, Peter (1983). The causal Theory of Reference. In: *Philosophical Papers* (vol. 1). Oxford: Oxford University Press, 2006, p. 117-160.
- WALTON, Kendall L. (1990). *Mimesis as Make-Believe*. Cambridge: Harvard University Press.
- WILLIAMSON, Timothy. (1994). *Vagueness*. Routledge.

Filosofia da Mente

Samuel C. Bellini-Leite

Compreender a mentalidade é atualmente uma das dificuldades filosóficas e científicas mais polêmicas, com pouco consenso envolvido, e principalmente, poucas propostas satisfatórias. Os principais problemas incluem o de dizer como a mente pode ter poder causal na matéria (causação mental – veja Amaral, 2001; KIM, 1998), o problema de determinar a individualidade de uma pessoa portadora de uma mente (problema da identidade pessoal – veja Coelho, 2020), o problema de determinar o conteúdo que a mente carrega (problema do conteúdo mental – veja De Menezes, 2020.), e como esse conteúdo porta um significado que faz referência ao mundo (problema do significado e da intencionalidade – veja Leclerc, 2018), dentre outros. Podemos dizer que todos esses estão de certa forma relacionados a um problema geral, conhecido como o problema mente-corpo (veja Fodor, 1981). A resposta para esse problema deve explicar a natureza da mente, e, havendo diferença em relação ao corpo, como esses dois se relacionam.

O problema da mente perpassa toda a história da filosofia e envolve diversas áreas de estudo, como a epistemologia, a metafísica, a lógica, a linguagem e a ética. Sendo assim, é comum encontrar em qualquer filósofo, teses que hoje podemos relacionar ao campo da filosofia da mente. Um filósofo historicamente lembrado pelo início de questionamentos especificamente mentais é o francês René Descartes (1641). Entretanto, é comum apontar o início formal da filosofia da mente ao início do século XX, a partir de interesses da filosofia analítica em compreender a relação de conceitos e proposições com o mundo (RUSSELL, 1921). Inclusive, esses dois pontos estão relacionados. A filosofia da mente, propriamente dita, surgiu de críticas à René Descartes feita por filósofos analíticos do século XX que começavam a se interessar pelo tópico (RYLE, 1949).

1. Críticas ao Dualismo de Substância

A proposta desenvolvida por Descartes é conhecida hoje como dualismo de substância, ou ao menos, consideram este um dos principais responsáveis pela elaboração e defesa dessa forma específica de dualismo (MASLIN, 2007). Essa visão de mente pode ser compreendida como parte do que Gilbert Ryle (1949) chamou de “doutrina oficial”. O dualismo de substância propõe que a mente é uma substância diferente do corpo, que possui características intrinsecamente distintas. Em particular, enquanto o corpo tem extensão, ocupando espaço físico e podendo ser dividido em pedaços, Descartes argumentou que a mente, ou a alma, é uma substância imaterial e sem extensão. Portanto, esta não ocuparia espaço físico e nem poderia ser dívida, característica que garantiria inclusive a imortalidade da alma após a morte do corpo. Se a alma é uma substância diferente do corpo que possui características distintas e que não podem ser divididas, então, pelo raciocínio, ela poderia continuar íntegra após o desmanche do corpo em pedaços (veja Descartes, 1641; e Maslin, 2007).

Evidentemente, a proposta de Descartes (1641) é mais complexa e elaborada do que esse parágrafo descrito acima, e o leitor fica convidado a se dirigir ao trabalho original. Entretanto, a filosofia da mente, propriamente dita, não se inicia em Descartes, mas na crítica ao dualismo de substância. Passaremos agora, então, para a explicação de críticas ao dualismo de substância, que convencem filósofos, desde o início do século XX, que outra resposta precisa ser dada ao problema mente-corpo.

O primeiro problema é que o dualismo de substância desconsidera a biologia ao tratar a mente como um fenômeno “tudo-ou-nada”. É intuitivo, em nossa cultura, pensarmos em seres humanos com almas, mas e se fôssemos pensar em animais com almas? Inicialmente não parece má ideia, por exemplo, um cachorro ter alma, se nós também temos. Entretanto, se a gente estende esse conceito na linha evolutiva, ele começa a ficar bastante estranho. Se a gente supuser que os animais possuem almas, teríamos que dizer que formigas, peixes, lesmas, estrelas do mar, dentre outros seres do reino animalia, possuem uma substância imaterial que de alguma forma acopla aos seus corpos durante a vida, mas perduram de uma forma não-física após a morte.

Descartes, inclusive, considerava animais como meras máquinas que

reagem ao mundo de uma forma quase que automática, ou seja, eles não possuiriam mentalidade, seriam corpos vazios de almas, mais próximos aos robôs (ou autômatos) do que seres humanos. Entretanto, atualmente, as evidências da biologia comparativa, do estudo de células e tecidos, da neurociência comparativa, e etologia comportamental, mostram que animais possuem uma consciência similar à nossa em vários aspectos (EDELMAN, BAARS e SETH, 2005; EDELMAN e SETH, 2009). Recentemente, um grupo de neurocientistas e estudiosos da consciência se reuniram para assinar uma declaração sobre a consciência de animais, a partir dessas evidências (LOW et al., 2012). Logo, esta visão dos animais enquanto autômatos foi cientificamente contestada. Assim, qualquer mentalidade que aplicamos aos seres humanos, precisamos aplicar também, pelo menos em parte, à diversos outros animais.

Entretanto, podemos, é claro, traçar um corte evolutivo e supor, por exemplo, que apenas mamíferos possuem alma. Mas assim não estaríamos coerentes com a noção de mudança gradativa da biologia evolutiva. Isto é, as mudanças na história da biologia ocorreram de forma gradativa durante a evolução e a mentalidade precisa seguir também este princípio. Além disso, o conceito começa a perder o sentido; poderíamos nos questionar como seria uma alma de uma ratazana, ou se morcegos teriam almas. O conceito de alma parece ter uma conotação moral, de individualidade e justiça. Por exemplo, culturalmente tendem a dizer que as ações dos seres humanos durante a vida influenciam após a morte de uma forma moral; se alguém for justo, ele terá benefícios, caso contrário terá prejuízos. É difícil pensar em como um morcego entraria nessa história moral. Por diversas razões, limitar a alma para os mamíferos ainda assim seria um corte bastante contraintuitivo.

Outro problema para o dualismo de substância é explicar a causação mental. Enquanto é possível concordar que explicar a causação mental é uma dificuldade para qualquer tese que proponha falar de mente, este problema é especialmente fatal para o dualismo de substância. Quando um teórico assume que existem estados mentais, é necessário que ele seja capaz de explicar como os estados mentais influenciam o corpo. Por exemplo, quando eu *penso* (estado mental: pensar) em mover meu braço e *decido* (estado mental: decidir) agir, como podem estes estados gerarem um movimento físico no meu corpo? Bem, pelo dualismo de substância, o pensamento é responsabilidade da alma, uma substância imaterial, sem extensão, não-física. Em contrapartida, uma interação

causal depende de mudanças físicas. Quando movimentamos uma peça de xadrez, estamos usando nossas mãos para aplicar força física ao objeto e deslocá-lo para a posição desejada. É difícil compreender como a alma, responsável pelo desejo e intenção, poderia desencadear correntes elétricas e químicas no cérebro que levam este comando para o braço e para as mãos.

De uma forma geral, é difícil de entender como haveria comunicação entre a alma e as partes do cérebro (como o córtex motor) que as evidências neurocientíficas mostram estar relacionadas ao movimento e à intenção do movimento. Descartes atribuiu à glândula pineal, uma estrutura pequena e única no cérebro, a responsabilidade de interação com a alma. Entretanto, além de não estar relacionada às funções de alto-nível como o pensamento (usualmente atribuídas ao córtex pré-frontal), simplesmente apontar uma estrutura física de comunicação não resolve o problema. Ainda precisaríamos saber como a alma (não-física) interage com a glândula pineal (física). Esse problema é especialmente fatal para o dualismo de substância pois, mesmo se supormos algum fenômeno que gera essa interação, uma força vinda de uma estrutura não-física mudaria todo o fechamento causal da física. Para que haja uma alma capaz de causar, precisamos, em essência, abandonar o que a física nos mostrou sobre a como funciona a estrutura do universo (mas veja Melo e Pessoa, 2016). Em termos morais e epistemológicos, uma pessoa que acredita em alma precisa dizer ao físico que ela compreende a estrutura causal do universo de uma forma que o físico não compreende, o que certamente não parece provável. Precisamos, na filosofia, ter respeito com as ciências e com os especialistas, não podemos formar preceitos que vão contra teses bem construídas da ciência, sem termos bons motivos para isso. No mínimo, enquanto podemos dizer que a filosofia fornece uma liberdade de explorar possibilidades que outras áreas acadêmicas geralmente não fornecem, podemos também dizer que quanto mais esses preceitos se distanciam da ciência, mais eles perdem a relevância para a comunidade acadêmica.

O fato de a intenção do movimento poder ser mapeada ao córtex, e não apenas o início causal da corrente elétrica relacionada ao movimento (HAGGARD, 2005), nos lembra de mais um problema para o dualismo de substância: se as funções como intenção, personalidade, memória, identidade pessoal, pensamento, desejo, podem ser funções do cérebro, o que restaria para ser função da alma? O que o dualismo de substância nos traz é um choque de

funções da alma e do cérebro, e se torna difícil determinar o que seria responsabilidade de cada substância.

Há um caso, clássico na história da neurociência, de um homem que teve o cérebro perfurado em um acidente de trabalho, mas sobreviveu, Phineas Gage (veja Damasio, 1996). Relatos da época dizem que houve uma mudança drástica em sua personalidade e seu caráter. Na realidade, como este é um caso antigo, existem diversos problemas relacionados à sua interpretação e à sua validade. De toda forma, atualmente existem inúmeros exemplos de alterações de personalidade, memória, identidade pessoal, assim como mudanças contraintuitivas na mentalidade das pessoas que sofrem lesões ou degenerações cerebrais, que mostram como essas funções necessariamente precisam estar relacionadas ao cérebro durante a vida (SACKS, 1985).

Diante dessa divisão de funções entre cerebrais e aquelas da alma, o teórico dualista precisa decidir, não se sabe como, o que é exatamente função de cada substância. Como essas funções são alteradas através de lesões e degenerações cerebrais, parece natural concluir que na realidade todas essas funções que antes eram atribuídas à alma, são simplesmente funções cerebrais. Afinal, o cérebro já sabemos que existe.

Considere a hipótese de que a alma levaria consigo, após o perecer do corpo, algumas marcas de uma pessoa que lhe garanta sua individualidade, como a memória, pensamento e personalidade. Agora vamos considerar um caso de uma pessoa que sofreu de doenças naturais na infância, bullying na adolescência e por isso mudou sua forma de ser. Finalmente, aos trinta anos começou a apresentar sintomas dissociativos, como perda de identidade e personalidades múltiplas. Por fim, sofreu de Alzheimer quando idoso, fazendo com que perdesse sua memória e alterasse ainda mais sua personalidade. Após a morte, o que essa alma levaria consigo? As características genéticas sem alterações da infância? A forma como ela viveu na adolescência? Quais das personalidades durante seus episódios dissociativos? E por fim, levaria suas memórias intactas ou apenas o que restou após o Alzheimer?

É muito raro que tudo isso aconteça a uma pessoa, apesar de haver muita comorbidade na psicopatologia, mas, de toda forma, todas essas dificuldades acontecem com certa frequência e com muitas pessoas. Assim, essas dificuldades psicopatológicas nos mostram que a nossa memória, nossa personalidade, e até mesmo nossa identidade estão intimamente ligadas a um

momento da nossa vida, a um estado do nosso cérebro e do nosso corpo. Não faz sentido dizer que esses estados continuariam após a morte, pois mesmo na vida tendemos a perder esses estados. Eles são estados, e assim, não são infinitos, são passageiros, como tudo na biologia.

Há razões para acreditar que o dualismo de substância é uma explicação homuncular. Um homúnculo é um homem pequeno que imaginamos para problemas que não conseguimos resolver. Quando não se compreendia sobre a estrutura do esperma, por exemplo, imaginava-se que poderia haver um homúnculo escondido que iria desenvolver na gestação. Além dessa explicação ser falsa, ela também não é efetiva. O problema teórico com essa explicação é que se o homúnculo estivesse no esperma, precisaríamos ainda de uma explicação sobre como o homúnculo é criado. Logo, para explicar a gestação, postulavam um homúnculo, mas a entidade postulada em si também não era explicada, fazendo com que a necessidade de explicação simplesmente mudasse de foco do esperma para uma entidade postulada, o homúnculo.

O mesmo pode ser dito sobre teses do dualismo de substância. Se a resposta para a pergunta “como nós pensamos?” for “temos uma alma que pensa”, temos como consequência uma terceira pergunta “como a alma pensa?”; isto é, nada foi explicado; a explicação apenas se deslocou do cérebro para a alma, e alma parece ser ainda mais difícil de explicar do que o cérebro.

Vale notar que há uma forma de explicação homuncular válida. Por exemplo, termos como “grupos neurais”, “sistemas” e “representações” podem agir de forma homuncular em explicações. Entretanto, nesses casos, as entidades postuladas são mais simples do que aquelas explicadas, podem existir e podem ajudar a conectar dois níveis de explicação, como o celular e o funcional, por exemplo. Esta forma de explicação, conhecida como funcionalismo homuncular, é justificável e ainda possui adeptos (LYCAN, 1990; SANT’ANNA, 2012). Isto é, em suma, a dificuldade da explicação homuncular do dualismo é essencialmente que a entidade postulada é mais complexa, misteriosa, ou seja, mais difícil de explicar, e implausível de existir, do que o fenômeno a ser explicado em si.

Por fim, há um último problema com uma característica não-física da alma, a falta de extensão. Pelo dualismo de substância, a alma não ocupa lugar físico no espaço, ela não pode ser maior ou menor, larga ou delgada, ela não pode ser mensurada. Entretanto, quando teóricos do dualismo de substância

afirmam que cada pessoa tem a sua alma, eles estão sendo incoerentes. Se alma não tem extensão, então não seria possível contá-la, não poderíamos dizer que “João tem uma alma” e “Letícia tem outra alma”, pois ao fazer isso estaríamos dando a ela extensão. Uma entidade sem extensão não poderia ser individualizada dessa forma e muito menos ser atribuída a uma pessoa como se fosse uma coisa única e separável, isto vai contra a própria definição de alma. Há aqui então, um problema considerado muito grave na filosofia, uma teoria contraditória. Isso faz ela não apenas ser falsa sobre o nosso mundo, mas impossível de ser verdadeira em qualquer mundo possível (veja Maslin, 2007).

O problema de contar almas é difícil de perceber inicialmente, mas em geral, é muito fácil gerar outros problemas para a doutrina oficial, basta tentar. Por exemplo, se considerarmos a possibilidade de espíritos desincorporados que falam, poderíamos nos questionar como eles poderiam produzir sons, se sons são ondas produzidas na laringe, pelas cordas vocais, devido à passagem de ar dos pulmões. Poderíamos questionar como eles supostamente encostariam em algo, e ao mesmo tempo passariam por paredes. Enfim, se uma pessoa estiver disposta a se questionar, ela consegue se livrar de mitos culturais da doutrina oficial com facilidade (PENNYCOOK, et al. 2012). Tente você também desenvolver novas objeções.

Abandonar o dualismo de substância pode ser particularmente difícil para algumas pessoas, pois ele geralmente está associado às crenças religiosas, que podem ser preciosas para estas. Entretanto, muitas vezes, e isto é muito comum mesmo, para pensar de uma forma filosófica precisamos estar dispostos a largar e reconstruir nossas crenças mais profundas, que mais se associam com quem pensamos ser, antes de começar a estudar. Esse tipo de mudança também pode ser a mais valiosa para nossa vida intelectual, é essa abertura que permite que um filósofo avalie as teses por si, considerando as razões para elas serem verdadeiras ou boas teses, sem que elas tenham vínculo com nossa personalidade. Afinal nossa personalidade é individual, já teses sobre o mundo devem valer igualmente para todos, afinal o mundo é o mesmo.

O trabalho da filosofia da mente se inicia quando o filósofo percebe que o dualismo de substância está equivocado e se pergunta então como ele pode ser substituído. Isto é, quando se elimina o dualismo de substância, o problema não está resolvido, pelo contrário, a partir dessa dissolução é que o filósofo da mente precisa entrar para responder os problemas de como a mente funciona no

mundo.

2. Propostas Anti-Mentalistas

A primeira postura dos filósofos analíticos devido à derrota do dualismo de substância foi a de igualar o mentalismo ao dualismo. Se todo tipo de mentalismo for dualismo, então o mentalismo, por consequência, também estaria equivocado. Podemos chamar de mentalista, qualquer tese em filosofia da mente ou psicologia que utilize termos mentais em sua explicação de base, isto é, termos como acreditar, conhecer, compreender, raciocinar, sonhar, imaginar, tentar, querer, pretender, dentro outros. Podemos chamar de anti-mentalista qualquer tese que tenha uma proposta para se livrar desses termos, ou que tente evitar falar de estados mentais. Existem diversas maneiras que filósofos propuseram se livrarem desses termos: os teóricos das teses de identidade (SMART, 1959; PLACE, 1970), os eliminativistas (DENNETT, 1991a; CHURCHLAND, 1984; LEAL-TOLEDO, 2018), os ilusionistas (FRANKISH, 2016) e os behavioristas (RYLE, 1949; SKINNER, 1957).

Teóricos de identidade, como Smart (1959) e Place (1970), não acreditam ser possível eliminar o vocabulário de estados mentais em teses da filosofia da mente, mas se perguntam, será possível falar de eventos internos e processos privados, sem assumir uma posição dualista na qual sensações e imagens mentais formam uma categoria separada dos processos físicos, com os quais são correlacionados? A estratégia que traçam é de argumentar que esses estados são idênticos à estados cerebrais. Isto é, um estado mental, como sonhar, seria idêntico ao processo cerebral relacionado a sonhar, não haveria diferença a ser apontada.

Place (1970) não acredita que seja possível derrubar sua tese em termos lógicos, afirma que tentativas de demonstrar logicamente a falsidade da afirmação “a consciência é um processo no cérebro” falham na diferenciação de tipos de significado da palavra “é”. O “é” de definição caracteriza uma junção de sujeito e predicado que torna a afirmação verdadeira por definição: um quadrado é equilátero; vermelho é uma cor; entender uma instrução é ser capaz de agir apropriadamente nas circunstâncias apropriadas. Já em afirmações que utilizam o “é” de composição, o sujeito e o predicado da frase não possuem uma relação de necessidade, e a veracidade precisa ser verificada empiricamente: sua mesa

é uma caixa velha, o seu chapéu é um conjunto de palha amarrado por um fio, uma nuvem é uma massa de pequenas partículas em suspensão no ar. Para Place (1970) a afirmação “a consciência é um processo no cérebro” está usando o “é” de composição, o que faz com que sua hipótese seja uma afirmação empírica. Ele acredita que seria preciso investigar empiricamente para descobrir se sua teoria da identidade é verdadeira ou falsa.

A teoria da identidade tem ainda duas vertentes, a versão tipo-tipo e a versão token-token. Na versão token-token, cada evento cerebral único é idêntico a um estado mental único, não haveria como formar categorias de identidade. Dessa forma, quando eu penso em um gato, há um evento cerebral ocorrendo que é idêntico a esse meu pensamento de gato, e apenas esse evento aqui e agora forma essa identidade. Por outro lado, na versão tipo-tipo, poderíamos dizer que ver é idêntico a ter certos estados do lobo occipital do córtex, em diversas pessoas. A identidade seria igual em diversos cérebros, sendo possível formar categorias. Esta versão supera a teoria token-token pois por ela seria possível ter um base para formar uma ciência, pois existe a possibilidade de encontrar verdades generalizáveis para diversas pessoas. Enquanto na teoria de token-token, se cada estado mental específico corresponde a um estado cerebral específico de um dado momento, não seria possível ter uma ciência que faça um estudo da ligação de estados de identidade, pois eles seriam únicos a cada ocorrência. Por outro lado, parece simplesmente falso que, quando dois sujeitos pensam em gato, os mesmos estados neurais estariam ativados. Assim, a teoria de identidade tipo-tipo é empiricamente falsa, pelo menos para diversas funções cerebrais que não possuem correspondências entre indivíduos.

Um problema geral dessa vertente de identidade é que em nosso melhor sentido do termo identidade, um conceito é idêntico a outro quando ele carrega todas suas propriedades. Mas um elefante rosa em nossa imaginação possui propriedades que não se aplicam aos neurônios que o produzem, os neurônios não são rosas. Assim, parece aceitável dizer que neurônios produzem a imaginação, mas não parece ser coerente dizer que nossa imaginação é idêntica aos neurônios, pois possuem propriedades distintas.

A corrente eliminativista (CHURCHLAND, 1984) desistiu de tentar mapear estados mentais em estados cerebrais, seja por tipo-tipo ou token-token. O que eles propõem é uma completa eliminação do uso de termos mentais em

qualquer explicação séria sobre a mente. Em vez de falarmos de dores, após estudos neurocientíficos descobrirem os estados cerebrais relacionados à dor, digamos a “Fibra C”, por mais que em nosso diálogo cotidiano usemos termos de estados mentais, a ciência pelo menos, deveria falar apenas dos estados cerebrais. Por exemplo, o cientista nãoalaria “ele está sentindo dor”, mas diria “estão ocorrendo fortes ativações de sua Fibra C”. Os eliminativistas acreditam que estados mentais são uma invenção da psicologia popular, a teoria cultural dos estados mentais. Eliminativistas variam sobre a utilidade da psicologia popular. Alguns acreditam que a psicologia popular apenas atrapalha em qualquer comunicação humana. Outros acreditam que ela foi útil evolutivamente e continua sendo útil culturalmente. Mas todos concordam que para a ciência, especificamente, esses termos precisam ser abandonados, como outros termos anticientíficos já foram no passado, como o *flogisto*, e evidentemente, a alma.

Uma versão atual das teses eliminativistas é o ilusionismo (Frankish, 2017). Esta tese difere termos mentais de estados qualitativos da consciência, como *qualia* (veja Nagel, 1974). Enquanto ela propõe que podemos falar de estados mentais como acreditar, pensar e desejar, ela não aceita a realidade de estados fenomenológicos; acredita que a introspecção é uma ilusão, apenas mais um viés cognitivo que sofremos. Entretanto, como ela não foca necessariamente em eliminar estados mentais básicos, ela não traz o conflito que as demais teses anti-mentalistas trazem para o atual capítulo.

Estas teses mencionadas acima ainda possuem um grande papel na filosofia da mente, mas podemos considerar o behaviorismo a maior corrente, em termos históricos, das teses anti-mentalistas. O behaviorismo ocorreu não apenas na filosofia com Bertrand Russell (1921), ou Ryle (1949), mas também na psicologia com Watson (1913) e Skinner (1957). Alguns teóricos behavioristas acreditam não haver uma mente para ser explicada, outros acreditam nos estados mentais, mas pensam que metodologicamente precisam ser ignorados para que haja uma ciência possível, e outros afirmam que conseguiremos compreender a mente melhor pelo comportamento do que pela introspecção. Apesar de haver diferenças entre todas essas propostas mencionadas, podemos dizer que todas são simpáticas, ao que Jerry Fodor (crítico do behaviorismo) chama de a tese essencial do behaviorismo.

Fodor (1968, p.51) chama de behaviorista qualquer doutrina que aceite a proposição ‘P’, onde a proposição P diz que para cada predicado mental que

pode ser colocado em uma explicação psicológica, precisa haver pelo menos uma descrição do comportamento com a qual ele tem uma conexão. Em contrapartida, qualquer doutrina que nega P pode ser chamada de mentalista. Fodor, diferentemente dos anti-mentalistas, diferencia o mentalismo do dualismo, onde o dualismo é qualquer doutrina em que mentes são compostas de uma substância irreduzível ao corpo. O dualista precisa ser um mentalista, mas um mentalista não precisa ser um dualista. Para o mentalista, basta afirmar que predicados mentais e predicados sobre o comportamento são logicamente distintos e ainda assim se aplicam sobre um mesmo e único corpo. Fodor (1968) argumenta que teóricos do início do século XX, quando discutindo entre o behaviorismo e o dualismo não perceberam que ambos poderiam estar errados. Fodor (1968) assegura ainda que enquanto argumentos de eliminação funcionam contra a ideia de múltiplas substâncias, eles não funcionam contra o mentalismo.

O behaviorista moderado não tenta eliminar os predicados mentais, ele meramente julga que esses predicados precisam ter um correspondente no comportamento. Fodor (1968) explica que um dos argumentos adotados pelo behaviorista é que nós conseguimos compreender os outros pelo comportamento, e o mentalismo pode nos levar à privação do conhecimento dos estados mentais, ou seja, à tese de que não podemos compreender os outros pelo comportamento. Como parece claro que sabemos quando alguém sente dor ou está feliz, então o mentalismo estaria nos levando para um problema que não condiz com a forma como compreendemos os outros.

Entretanto, Fodor (1968) argumenta que a ligação entre predicados mentais e predicados comportamentais não é necessária, mas apenas contingente, e não anômala (veja Fodor, 1989), por seguir regras *ceteris paribus*, ou seja, regras que ‘geralmente ocorrem se tudo estiver dentro do esperado’. Isto é, o tipo de regra entre comportamentos e predicados mentais que temos possuem a seguinte forma “se alguém está gritando, batendo o pé, e socando uma parede ele estará, *ceteris paribus*, nervoso”. Mas essa conexão *ceteris paribus* entre os predicados comportamentais e os predicados mentais não garante a tese P que o behaviorista precisa. Pelo contrário, essa conexão fraca parece garantir algo muito mais próximo ao mentalismo, pois haverá casos onde alguém grita, bate o pé e soca e mesmo assim não está nervoso, inclusive a pessoa pode estar nervosa e exibir sorrisos. Lembrando que, o mentalista

evidentemente concorda que muitas vezes conseguimos deduzir os estados mentais nos baseando no comportamento.

O psicólogo, que, com razão, preocupa-se em estudar a mente de forma científica, aplica o behaviorismo como uma ferramenta metodológica para supostamente compreender os fatos observáveis de uma forma neutra e eficaz. Para contrapor a isso, Fodor (1968) considera um caso em que a resposta B é condicionada ao estímulo A e uma resposta C ao evento B. É de se esperar que a apresentação de A elicite C mesmo que o par A-C nunca tenha sido condicionado. A tendência é supor que B tem um papel ocorrente mediador e não observável. Fodor (1968) considera plausível que B seja produzido de uma forma não observável e influencie a resposta C. Poderíamos ainda supor essa ocorrência com mais segurança se o par A-C gastar mais tempo para ocorrer do que outro par semelhante que não pareça ter um fator interno mediador. Fodor (1968) argumenta que se pesquisas mostrarem que a melhor hipótese para explicar o comportamento depende de eventos internos mediadores, sem um comportamento observável correspondente, o psicólogo teria que escolher entre abandonar a metodologia original ou abandonar a melhor explicação dos fatos. Considerando que nunca devemos abandonar a melhor forma de compreender os fatos, e que a metodologia inicial foi proposta justamente para cumprir o objetivo de explicar os fatos, parece que a solução mais aceitável é de abandonar o princípio metodológico.

Além das críticas específicas mencionadas para cada forma de anti-mentalismo, há uma crítica que se aplica, de forma fatal, a todas as teses anti-mentalistas, a permanência das proposições de identidade.

Mesmo supondo alguma forma de identidade (tipo-tipo ou token-token) entre estados mentais e físicos, as proposições que versam sobre essa identidade seriam contingentes e passíveis de análise. Então, para essa identidade existir, ainda assim, precisamos concordar que existem essas proposições de identidade que não podem ser formuladas sem o uso da linguagem mentalista. Por exemplo, a proposição “crenças são estados neurais X” inclui um termo mentalista, a saber: crenças. Se a ciência começar a usar o termo “estados X” em vez de crença, isto não muda o fato de haver essa ligação inicial baseada na proposição de identidade, que contém o termo crença e que, então, faz a referência a um estado mental. Em contraste, você pode eliminar, por exemplo, o termo “alma” da linguagem de identidade e não haverá perda

explicativa para a ciência.

O eliminativista pode acreditar se safar dessa crítica. Entretanto, estudando a metodologia da neurociência de forma séria, percebe-se que não há como fugir. Para ser possível um neurocientista atribuir uma função a uma região do cérebro, ele precisa pensar sobre os estados mentais, e conversar sobre os estados mentais com o sujeito da pesquisa. Para estudar a imaginação, ele fala para o sujeito do seu experimento, por exemplo “imagine um elefante rosa”. Então, nesse caso, não apenas ele usou seus próprios estados mentais para definir seu objeto de estudo em termos mentais – a imaginação, como atribuiu estados mentais ao sujeito do experimento utilizando termos mentais na comunicação, para apenas depois dizer que certa área do cérebro “Y” corresponde à imaginação. Posteriormente, dizer que não devemos falar de imaginação, e apenas devemos falar de Y, não tem valor algum real de eliminação, pois tudo que estaríamos fazendo é atribuindo um código Y (uma área cerebral específica, um neurotransmissor, uma reação química) para falar do estado mental imaginação, estaríamos, em essência, apenas mudando o nome da palavra, mas a referência continua sendo o estado mental. Conseguiríamos compreender o significado de “Y” pois sabemos, pela introspecção e pela psicologia popular o que é imaginação. Logo, até mesmo em teses eliminativistas, a crítica da permanência das proposições de identidade de Fodor (1958) se aplica de forma fatal.

Vale notar que apesar dos comentários acima, teses anti-mentalistas ainda formam uma corrente comum e aceita na filosofia da mente (veja Leal-Toledo, 2018).

3. Mentalismo de Processamento de Informação

As críticas que o behaviorismo sofreu (CHOMSKY, 1959; FODOR, 1968) abriram espaço para novas formas de falar sobre a mente. Como os predicados mentais precisam necessariamente existir na psicologia, o dilema disponível não é entre usar ou não usar predicados mentais, mas sim entre ter uma linguagem popular dos estados mentais (ideias, intenções, crenças, desejos) ou uma linguagem científica destes. A única corrente em toda a história da filosofia e da psicologia que forneceu uma linguagem científica para tratar dos estados mentais foi a ciência cognitiva, através da hipótese do processamento de

informação. Essa hipótese trouxe o mentalismo para destaque no estudo da mente.

A hipótese do processamento de informação diz que estados mentais são processos informacionais ocorrendo no cérebro humano, de forma semelhante a como um computador (material) roda um software (padrões de energia). Não se trata apenas de uma analogia. Apesar de nenhum filósofo acreditar que o cérebro é como um computador clássico inventado no século XX, uma grande parte concorda (FODOR, 1975; DENNETT, 1991a; CHALMERS, 1995) com a afirmação metafísica de que a mente é um processador de informação.

A principal mudança tecnológica que influenciou o início dessa vertente foi o surgimento de computadores, em especial as aplicações da inteligência artificial. Não apenas em termos de aplicações, mas também em termos filosóficos. Alan Turing, um matemático que auxiliou a decodificar códigos nazistas na segunda guerra mundial, propôs uma das primeiras formulações matemáticas de um computador (TURING, 1936) e também influenciou os estudos da mente através de questionamentos sobre como o pensamento está relacionado ao processamento de informação de computadores (TURING, 1950).

Uma máquina de Turing (1936) é uma abstração que, de forma geral, consiste em uma fita (infinita) dividida em quadrados onde um símbolo pode ser escrito; um leitor que considera apenas um quadrado de cada vez; e um conjunto de regras que funciona como uma tabela de instruções para o que a máquina faz. O leitor examina o símbolo na fita e prossegue conforme a tabela de instruções determina. A máquina, a qualquer momento, pode substituir algum símbolo na fita, mover o leitor para a esquerda ou para a direita ou parar. A partir da leitura de instruções e manipulação de símbolos, em princípio, essa máquina pode praticar qualquer tipo de computação. Esta máquina é uma abstração matemática. No entanto, a arquitetura Von Neumann foi planejada para instanciar fisicamente uma versão dessa máquina.

A arquitetura de Von Neumann (Von Neumann, 1945) é um projeto para implementar um computador digital. Consiste em uma unidade de processamento, uma unidade de controle, armazenamento primário e mecanismos de entrada/saída. A unidade de processamento contém uma unidade aritmética/lógica que executa instruções aplicando procedimentos aritméticos ou lógicos em um código binário e um registro que armazena

rapidamente etapas de tais operações. Este sistema é análogo ao que lê e aplica as alterações na fita no Máquina de Turing. Uma unidade de controle contém um registro de instruções, que por sua vez contém as instruções a serem processadas em um determinado momento, mediando o fluxo de processamento, fornecendo sinais de tempo e controle. É isso que torna a ideia de uma tabela de instruções fisicamente possível. Tais instruções são mantidas no armazenamento primário, juntamente com os dados que foram trabalhados na unidade de processamento. Portanto, trata-se de uma memória que é análoga à fita na Máquina de Turing, mas que também armazena as instruções da tabela de instruções. Mecanismos de entrada e saída são o que permitem comunicação entre este núcleo central com todas as suas outras partes, como seu disco rígido externo.

Bem, podemos ver que a implementação de um computador digital tem uma estreita semelhança arquitetônica com uma Máquina de Turing, e isso não seria verdade para os casos de um computador quântico, um computador funcionando em paralelo ou um cérebro. De toda forma, os primeiros proponentes da hipótese de processamento de informação não possuíam ligação com a neurociência e acreditavam ser possível falar de uma mente, de certa forma, independentemente de sua implementação neural.

Newell e Simon (1976) e Newell (1980) propuseram uma primeira base para se pensar como agentes dotados de um sistema de processamento de informação poderiam resolver problemas no mundo real, elaborando a hipótese do sistema de símbolos físicos.

Newell e Simon (1976, p.116) afirmam que “um sistema de símbolos físicos consiste em um conjunto de entidades, os símbolos, que são padrões físicos (por exemplo, elétricos) que podem ocorrer como componentes de outro tipo de entidade, chamada de expressão (ou estrutura de símbolo).” Operações manipulam tais símbolos, criando, copiando, editando e destruindo expressões. Ao contrário da Máquina de Turing, para um sistema de símbolos físico é essencial que este esteja em um mundo dos objetos. Isso porque um dos papéis fundamentais dos símbolos e expressões é o de designar objetos. Newell e Simon (1976, p.116) explicam que “uma expressão designa um objeto se, dada a expressão, o sistema pode afetar o próprio objeto ou se comportar de maneiras dependentes do objeto” e, assim, ganhar “acesso” ao objeto, via expressão. Como Newell (1980, p.156) coloca, “ter X (o símbolo) é equivalente a ter Y (a coisa

designada) para a finalidade do processo P". Qualquer objeto pode ser designado por meio de símbolos e expressões. Portanto, Newell (1980) afirma que os números em uma máquina de Turing não são símbolos como no caso de sistemas de símbolos físicos. Isso porque, enquanto os dígitos de uma máquina de Turing funcionam como *tokens* de expressões, elas não são, no entanto, acopladas a alguma estrutura de dados.

Ter símbolos capazes de designar discretamente concede ao sistema algumas propriedades, como transitividade. Se X é um símbolo que designa Y e Y designa Z, então X designa Z. Agora, só obtemos isso se X receber um valor fixo e discreto, assim, não pode haver incerteza sobre o valor de um símbolo, como se teria sobre uma variável. Em outras palavras, um símbolo designa um único objeto. Essas propriedades se estendem para produtividade e composicionalidade (veja Fodor e Pylyshyn, 1988).

Se uma expressão designar um processo do sistema e se o sistema puder executar tais processos, diz-se que o sistema pode interpretar a expressão. Outra maneira de entender isso é que expressões têm algum tipo de significado para a máquina quando ações (ou processos) podem ser derivadas delas. Tal processo é ativado pelo controle (uma parte do sistema de símbolos). O que o controle faz é reunir, em um determinado momento, dados e operadores. Interpretar expressões e aplicar operadores é o que o sistema de símbolos físicos faz para a produção de novas expressões. Nas palavras de Newell (1980, p.159) "esta organização implica um requisito para a memória de trabalho no controle para gravar os símbolos para o operador [...]" e também que "[...] a memória de trabalho é uma característica invariante dos sistemas de símbolos". Vale notar que um dos constructos da psicologia cognitiva com maior base de evidência é a memória de trabalho (BADDELEY e HITCH, 1974) que possui essas características bastante semelhantes àquelas mencionadas por Newell (1980).

O sistema por si só não necessariamente possui habilidades para resolver problemas. Assim, Newell e Simon (1976, p.120) explicam a resolução de problemas através da hipótese de busca heurística: "as soluções para problemas são representadas como estruturas de símbolos. Um sistema de símbolos físicos exercita sua inteligência no problema, resolvendo pela busca - isto é, gerando e modificando progressivamente as estruturas simbólicas até produzir uma estrutura de solução."

Como Newell e Simon (1976) explicam, para esses sistemas, os problemas surgem de uma forma específica. Para que haja um problema, deve haver um espaço de problema ou um espaço de estados. Esse espaço inclui todos os passos possíveis que o processo pode dar ao resolver o problema, incluindo o estado inicial, várias soluções e etapas intermediárias. Também deve haver um teste que indica o estado que a máquina deve estar para atingir um objetivo. E, finalmente, deve haver geradores de movimento, que são processos de modificação de uma situação no espaço de problema. Newell e Simon (1976, p.121) explicam que “a tarefa que um sistema de símbolos enfrenta, então, quando é apresentado a um problema e um espaço de problema, é usar seus recursos de processamento limitados para gerar possíveis soluções, uma após a outra, até encontrar uma que satisfaça o teste de definição de problemas”. O sistema precisa, entretanto, se organizar para que as soluções tenham uma alta probabilidade de ocorrerem. Um sistema de símbolos exibiria inteligência para um dado problema na medida em que mostra que tem alta probabilidade de encontrar soluções para o problema, em vez de caminhar aleatoriamente ou por tentativa e erro em um espaço de estados.

Assim, no caso de sistemas de símbolos físicos, um espaço de problema contém todas as formas de expressão que poderiam ser dadas à forma inicial considerando os geradores de movimento disponíveis. Assim, cada expressão obtida no processo é uma modificação da expressão inicial. Além disso, os geradores de movimento incorporam regras, como a lógica, aritmética ou outra. Estas regras contêm diretrizes gerais de resolução de problemas e outras especificidades. Por exemplo, uma regra geral seria algo como: “não produza contradições”. Uma específica estaria relacionada ao problema em questão, como “crie a menor expressão possível”. Em qualquer etapa, as modificações não são caóticas, mas sempre dependem dessas regras incorporadas pelos geradores de movimento e o estado atual do processo, um dado local no espaço do problema.

Pense na resolução de problemas de um sistema cognitivo como um caminhar sobre as diversas possibilidades de resposta a um problema. Os passos tomados em um espaço de problema formam ramificações de uma árvore. Se todos os estados (ou a maioria dos estados) de um espaço de problema foram visitados (pesquisa de força bruta), o processo incorreria em explosão computacional (não haveria tempo para ocorrer na realidade), a menos

que o espaço do problema fosse muito pequeno (o que não ocorre em problemas relevantes para o mundo). Assim, a busca deve ser heurística no sentido de que deve tomar o menor número de etapas possíveis para obter soluções.

Como Newell e Simon (1976, p.123-124) afirmam: “Nossa análise de inteligência igualou-a à capacidade de extrair e usar informações sobre a estrutura do espaço do problema, de modo a permitir que uma solução de problema seja gerada rápida e diretamente quanto possível” e “a tarefa da inteligência, então, é evitar a ameaça sempre presente da explosão exponencial de busca.”

Fodor (1975) propôs uma tese bastante popular de que nosso cérebro instancia algo como um sistema de símbolos físico de Newell (1980), o que ele chamou de linguagem do pensamento. Fodor (1975) argumenta ainda que essa linguagem é mais básica que a linguagem natural (português ou inglês) e que seria impossível aprender qualquer língua sem antes ter a linguagem do pensamento. Esses desenvolvimentos formam o que hoje é chamado de teoria computacional (ou representacional) da mente (FODOR, 1985). Interessantemente, essa proposta insere os estados mentais como algo natural do cérebro enquanto processador de informações, há um realismo dos estados mentais. Os estados mentais seriam, assim, *padrões*¹²⁵ elétricos, ou químicos, que formam símbolos que designam objetos do mundo real, e simulam espaços de problemas para serem investigados por buscas heurísticas ou qualquer outra técnica computacional/informacional de resolução de problemas. O mentalismo, ganha com essa proposta, sua primeira hipótese ontológica ligada ao fisicalismo ou materialismo na história. Em outras palavras, a mente é inserida em uma ontologia física (ou informacional), sem a necessidade de substâncias postuladas, afinal computadores existem, e sem a necessidade de teses eliminativistas.

Historicamente, a principal crítica à hipótese do processamento de informação é a analogia do quarto chinês. Searle (1980) gera um experimento de pensamento no qual uma pessoa em um quarto recebe a tarefa de traduzir

¹²⁵ A palavra “padrões” está em itálico para enfatizar que a ontologia da mente nessa proposta não é elétrica ou química, a mente é compreendida como padrões de atividade de processos físicos.

chinês com base em regras procedimentais e uma lista de instruções que pode acessar. A pessoa recebe letras em chinês, segue as instruções de modificação e passa a resposta adiante. Searle (1980) argumenta que essa pessoa não saberia falar chinês, ela apenas estaria aplicando regras de uma forma cega, assim como qualquer processador de informação. Dessa forma, ele desenvolve o argumento de que processadores de informação não possuem acesso ao significado do conteúdo que processam, e como seres humanos possuem tal acesso, eles não poderiam ser processadores de informação.

Apesar da crítica do quarto chinês ser repetida diversas vezes na história da filosofia da mente, assim como na cultura popular, interessantemente, é pouco discutido como Daniel Dennett (1980; 1991b) apresenta respostas bastante convincentes a esse problema. Dennett explica como a metáfora do quarto chinês está equivocada em colocar o homem responsável pelas manipulações como o agente que sabe o significado da linguagem, isso faria da teoria do processamento de informação uma proposta homuncular no sentido ruim. Na realidade, o significado da linguagem emergiria, para a metáfora estar correta, de diversos quartos desses em interação (DENNETT, 1980), cada um com seu funcionamento específico (sistemas cognitivos variados), todos interligados e mapeados pela evolução, por seleção natural, para reagir de certa forma e significar coisas específicas (DENNETT, 1991b). Apesar de funcionar na base da sintaxe, agentes cognitivos (máquinas treinadas ou cérebros) transformam um certo símbolo em outro apenas quando estão em certa relação semântica (FODOR, 1985; mas veja FODOR e PYLYSHYN, 2015).

Há, entretanto, um problema similar ao problema do significado, como repetido por Searle, que afeta de forma mais séria a hipótese do processamento de informação, o problema do *frame*, ou o problema da relevância (DREYFUS, 1972; PYLYSHYN, 1987). Este problema se refere à dificuldade encontrada por um sistema que trabalha com expressões representando o mundo para agir de forma fluida. As tentativas iniciais da inteligência artificial representacionista falharam, pois os robôs não conseguiam agir, por terem de considerar e catalogar inúmeras mudanças irrelevantes que acontecem no mundo. Este problema atualmente vem sendo tratado através de propostas incorporadas (CLARK, 1997; BROENS e GONZALEZ, 2006.) e através de representações probabilísticas e preditivas (CLARK, 2016; BELLINI-LEITE, 2018). Ainda, em termos técnicos, houve avanço considerável, que relativizou o problema, por exemplo, com o

desenvolvimento de *machine learning*, especialmente o *deep learning*, que fazem máquinas realizarem atualmente funcionalidades que antes eram pensadas serem impossíveis para máquinas em geral, por causa do problema do *frame*. Isso não significa que o problema tenha sido superado, mas não se trata mais de um problema fatal para a hipótese do processamento de informação, considerando que mudanças de perspectivas filosóficas e tecnológicas têm avançado em sua solução.

Em contrapartida, um problema fundamental não abarcado pela hipótese de processamento de informação é o problema dos sentimentos e da experiência. David Chalmers (1995) é um filósofo contemporâneo que aceita o mentalismo de processamento de informação, mas acredita que a mente, mais especificamente a consciência, não está completamente abarcada por essa tese. Enquanto a mente compreende estados inconscientes, como o plano motor, que permite que andemos de bicicleta sem prestar atenção nos detalhes do pedal, os quais poderiam ser explicados por processamento de informação, há também uma parte da mente, a consciência, que possui estados qualitativos, uma experiência de estar consciente (NAGEL, 1974).

O problema da experiência como elaborado por Chalmers (1995) diz que existem dois problemas da consciência, o problema fácil e o problema difícil. O problema fácil da consciência é aquele que pode ser respondido em termos de comportamento, fisiologia e teorias da cognição. Já o problema difícil, o problema da experiência, está em explicar como e por qual motivo uma experiência ocorre em conjunto com o comportamento e o processamento de informação. Da forma como a mente é descrita em termos científicos, por exemplo, pela neurociência cognitiva, não haveria motivo algum para a cognição vir acompanhada de experiência como dores, cores, sentimentos dentre outras. O problema pode ainda ser especificado em dois: (1) qual é a natureza da experiência (NAGEL, 1974); isto seria algo como encontrar uma forma de compreender onde o mundo fenomenológico se encontra em relação ao mundo objetivo; e, (2) ao explicar a natureza da experiência, deve-se ter também uma história sobre como essa experiência emerge do funcionamento cerebral, ou em geral, qual é a ligação da experiência com o comportamento e processamento de informação em seres vivos.

Chalmers (1996) tentou resolver o problema da experiência a colocando como algo fundamental na física, como são matéria e espaço, e também

cogitando o pampsiquismo, a tese de que todos os objetos possuem algum tipo de mentalidade. Entretanto, Chalmers (1995, 1996) foi reconhecido mais pela forma de caracterizar o problema do que pela resposta, pois sua resposta não trouxe consenso algum para a comunidade acadêmica.

Na realidade, existem inúmeras tentativas de resolver o problema da experiência como colocado por Chalmers (1995), aceitando o mentalismo e também acreditando haver a necessidade de maiores explicações sobre a natureza da experiência consciente. Podemos mencionar, por exemplo, o dualismo de propriedades e o monismo não redutivo com suas diferentes formas (veja Maslin, 2007), o monismo reflexivo (VELMANS, 2008), o monismo de triplo-aspecto (PEREIRA JR., 2015) e o naturalismo biológico (SEARLE, 1990).

Entretanto, ainda que essas propostas se mostrem preocupadas em explicar como pode haver diferentes níveis e como eles interagem, enquanto não houver uma explicação aceita de qual seria a natureza da experiência consciente, que não simplesmente redescreva a fenomenologia (BELLINI-LEITE, 2015), essas parecem meramente variações linguísticas; termos diferentes para ideias muito parecidas – e, enquanto proliferarem teorias, nenhuma será suficientemente consensual.

A verdade é que alguma explicação extra seria necessária para o mentalismo de processamento de informação resolver os problemas de Chalmers. Pense o caso dos sentimentos e das emoções. Não parece nem um pouco claro que robôs, que processam informação, terão sentimentos ou emoções, a não ser que adotemos posições behavioristas, eliminativistas ou ilusionistas sobre esses estados. Se aceitarmos a experiência como real, e a descrição fenomenológica como algo que ainda depende de mais explicações metafísicas, como acredita Chalmers (1996), então, intuitivamente, não temos motivo algum para acreditar que nossos computadores, se ficarem mais complexos, irão começar a ter sentimentos ou emoções. Esses aspectos parecem não serem apenas de processamento de informação, mas estarem ligados ao nosso corpo biológico, como argumenta Searle (1990) corretamente. Intuitivamente parece haver algo de especial na biologia, que gera estados de experiência em diversos seres vivos, o que não é o caso na robótica. Não parece ser plausível dizer que uma câmera de supermercado consegue ver, da mesma forma que um cachorro, nem mesmo se essa câmera estiver ligada a um sistema de inteligência artificial que processa a imagem com facilidade, pois, ainda assim

não haveria para a câmera a experiência de enxergar.

Logo, apesar de o mentalismo de processamento de informação ser a corrente melhor aceita na neurociência cognitiva (GAZZANIGA, IVRY e MANGUN, 2006) e estar ainda muito forte na filosofia da mente (DAVENPORT, 2012; COLTHEART, 2012), esta hipótese parece ainda em desenvolvimento. Ela ainda se desenvolve em relação à natureza das representações mentais, enquanto paralelas (CLARK, 1989), probabilísticas e preditivas (CLARK, 2016), incorporadas (CLARK, 1997), estendidas (CLARK e CHALMERS, 1998; COELHO, 2017), simbólicas (FODOR e PYLYSHYN 1988), além das propostas enativistas (CARVALHO, 2018; DI PAOLO e THOMPSON, 2014), na esperança de conseguir resolver os problema do significado (FODOR E PYLYSHYN, 2015), relevância (PHYLYSHYN, 1987) e abdução (FODOR, 2001; VITTI-RODRIGUES, MATULOVIC e GONZALEZ, 2017), e precisa de acréscimos à teoria que deem conta da experiência consciente (CHALMERS, 2004), para que não haja a necessidade de novas propostas eliminativistas ou behavioristas para esconder a nossa lacuna de conhecimento.

4. Conclusão

O leitor pode pensar que a história da filosofia da mente foi de alguma forma circular e que, como teses anti-mentalistas e mentalistas também são problemáticas, o dualismo de substância seria novamente um concorrente apto. Isto está equivocado, enquanto o dualismo é nobre no sentido de tentar responder questões da consciência, em vez de negá-las, o problema é que estas respostas estão fundamentalmente equivocadas. Isto é, enquanto o anti-mentalismo e o mentalismo contemporâneo possuem problemas a serem resolvidos – alguma versão dos dois pode estar correta¹²⁶ –, o dualismo de substância é o único que já foi suficiente e convincentemente demonstrado estar equivocado, tanto por argumentos quanto por evidências científicas. Portanto a negação do dualismo de substância continua sendo o ponto de partida da

¹²⁶ Por exemplo, a adoção do processamento de informação para explicar termos da psicologia popular como crenças e desejos, associado ao ilusionismo (FRANKISH, 2017) para responder ao problema da experiência, ou associado à uma proposta da virtualidade da experiência (REVONSUO, 1995) são caminhos bastante plausíveis.

filosofia da mente. O que passa a ser necessário é o desenvolvimento de novas teses que resistam aos problemas que ainda faltam ser enfrentados. Não é exagero afirmar que uma possível resolução desses problemas afetaria em grande parte diversas outras áreas da filosofia, como a epistemologia, a linguagem, a lógica, a ética e a metafísica. Este é um ponto de trabalho fortíssimo atualmente, que precisa ainda de muita pesquisa e apoio. Assim, o leitor está convidado a estudar o campo para além dessa introdução, para que possa no futuro desenvolver teses que contemplem as evidências, ao mesmo tempo sendo capaz de atingir os problemas descritos no capítulo, assim como outros que podem ser encontrados na literatura da filosofia da mente.

Questões de Revisão

1. O que é o dualismo de substância?
2. Relate como o dualismo de substância está ligado às crenças culturais.
3. Explique três críticas ao dualismo de substância.
4. Explique por qual motivo o dualismo de substância foi considerado superado mesmo considerando que todas as demais propostas também possuem problemas.
5. O que são termos mentais?
6. Qual é a diferença entre a teoria da identidade mente-cérebro e o eliminativismo?
7. Explique uma crítica de Fodor às teses anti-mentalistas.
8. Descreva uma dificuldade de igualar um estado mental à um comportamento.
9. Por que o mentalismo não precisa ser um dualismo de substâncias?
10. O que é um sistema de símbolos físicos de Newell?
11. Qual é a ontologia da mente na hipótese do processamento de informação?
12. O que ainda está em desenvolvimento na hipótese do processamento de informação?
13. Qual é o problema difícil da consciência de Chalmers?

Questões de Discussão

1. Sugira quais das críticas ao dualismo de substância você considera mais

- eficaz.
2. Por que ideias de alma ainda são prevalentes em nossa cultura?
 3. Como é a mente de um animal?
 4. Animais possuem qualidades do sentido que jamais conseguiremos ter?
 5. A sua mente é seu cérebro? Há diferença?
 6. Imagine um elefante rosa. Qual é a ontologia desta imagem?
 7. Como propriedades aparentemente não-físicas da mente podem interagir com o mundo físico-causal?
 8. Acredita que robôs possam um dia sentir dor e prazer? Por quê?
 9. Procure por exemplos de ilusão de cor na internet (como o vestido preto e azul ou branco e dourado; ou o tênis rosa ou verde) e responda, a cor é uma propriedade do objeto ou da mente humana? Existe alguma posição intermediária?
 10. Qual é a ontologia de um desejo?
 11. Tente definir o que são crenças.
 12. Há ainda um caminho plausível para teses anti-mentalistas?
 13. Quais são as semelhanças e as diferenças entre mentes e computadores?

Sugestão de leitura

A filosofia da mente (TEIXEIRA, 2016) é um tópico vasto que faz conexões com diversas ciências como a psicologia, inteligência artificial, linguística, biologia, neurociências, antropologia, física, dentre outras. Além disso, há uma pluralidade de visões e não há, muito claramente, uma ordem histórica de autores que deve ser seguida. O que pode ser feito, entretanto, é uma seleção de alguns problemas da filosofia da mente que o leitor se interessa para que busque conteúdo sobre esses problemas em diversos autores. Esse capítulo apresenta alguns desses problemas, mas não todos. Assim, seria interessante começar os estudos com um manual mais completo de introdução à filosofia da mente (MASLIN, 2007; LEAL-TOLEDO, GOUVEA e ALVES, 2018).

Tendo selecionado um problema, a sugestão é deixar o livro de introdução e obter conhecimentos mais específicos sobre o problema. Por exemplo, para complementar o atual capítulo seria interessante o leitor interessado ler os diálogos originais entre Searle (1980; 1990) e Dennett (1980; 1991b), para formar sua opinião sobre o quarto chinês (veja também Vargas,

2014, para uma comparação dos dois autores). O leitor pode ver as diferenças no tratamento da consciência através de propostas eliminativistas (CHURCHLAND, 1984; DENNETT, 1991a; LEAL-TOLEDO, 2018) e propostas que consideram a experiência qualitativa (CHALMERS, 1996; CHALMERS, 2004). Pode também, através de um mesmo autor, ver mudanças em relação a noção de representação e o problema do *frame* (CLARK, 1989, 1997, 2016).

Alternativamente, é importante verificar teses fundamentais e gerais que marcaram a história do estudo da mente, como o behaviorismo (RYLE, 1949; SKINNER, 1957; WATSON, 1913) e a versão simbólica do processamento de informação (NEWELL, 1980; FODOR, 1985; FODOR e PYLYSHYN, 1988), e eventualmente chegar em sugestões atuais sobre como enfrentar o problema da consciência (METZINGER, 2004; VELMANS, 2008; PEREIRA JR., 2015; FRANKISH, 2017).

Por fim, vale ressaltar que o recorte feito por esse capítulo não representa a única forma de fazer filosofia da mente, existe também uma corrente ligada à tradição fenomenológica (ABATH e CAMINHA, 2012; CAMPOS, 2017; NOË, 2004), trabalhos mais fortemente ligados à filosofia analítica (MENDONÇA, 2007; LEWIS, 1972) e também aqueles baseados na filosofia da linguagem (BARWISE e PERRY, 1981; PERINI-SANTOS, 2014; SEARLE, 1969).

Referências

- ABATH, A.; CAMINHA, I. Merleau-Ponty e o fisicalismo. *Revista de Filosofia Aurora*, v.24, n.35, p.615-638, 2012.
- AMARAL, F. Causação Mental: Onde Estivemos e Onde Estamos. *Psicologia: Teoria e Pesquisa*, Brasília, v.17, n.3, p.235-244, 2001.
- BADDELEY, A.; HITCH, G. Working Memory. *Psychology of Learning and Motivation*, v.8, p.47–89, 1974.
- BARWISE, J.; PERRY, J. Situations and attitudes. *The Journal of Philosophy*, v.78, n.11, p. 668-691, 1981.
- BROENS, M.C.; GONZALEZ, Maria Eunice Quilici. Um estudo do conhecimento não proposicional no contexto da teoria da cognição incorporada e situada. *Manuscrito (UNICAMP)*, v.9, p. 729-751, 2006.
- BELLINI-LEITE, S. Explicação metafísica, causação e interações neuro-astrocíticas: uma réplica a “o conceito de sentimento no monismo de triplo

- aspecto". *Kínesis-Revista de Estudos dos Pós-Graduandos em Filosofia*, v.7, n.14, 2015.
- BELLINI-LEITE, S. Surfing Uncertainty: predições situadas?. *Ciências & Cognição*, v.23, n.2, p.255-262, 2018.
- CAMPOS, V. Pondo A Mão Na Consciência (Incorporada): sobre o estatuto do corpo no pensamento de Jean-Paul Sartre. *ConTextura*, v.9, n.11, 2017.
- CARVALHO, L. A Realização da Mente: Crítica da Teoria Enativa ao Conceito de Propriedade Emergente. In: LEAL-TOLEDO, G. L.; GOUVEA, R. A. S.; ALVES, M. A. S. (Org.). *Debates Contemporâneos em Filosofia da Mente*. 1ed. São Paulo: FiloCzar, p. 113-129, 2018.
- CLARK, A; CHALMERS, D. The extended mind. *analysis*, v.58, n.1, p.7-19, 1998.
- CLARK, A. Connectionist minds. In: *Proceedings of the Aristotelian Society*. Aristotelian Society, Wiley, 1989. p. 83-102.
- CLARK, A. *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press, 1997.
- CLARK, A. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, 2016.
- CHALMERS, D. Facing up to the problem of consciousness. *Journal of Consciousness Studies*. v.2, n.3, 1995.
- CHALMERS, D. *The conscious mind: In search of a fundamental theory*. Oxford university press, 1996.
- CHALMERS, D. Como Podemos Construir Uma Ciência Da Consciência? (ALVES, M.; RAMOS, J. Tradução). *Prometheus-Journal of Philosophy*, n. 33, (Trabalho original publicado em 2004), 2020.
- CHOMSKY, N. A Review of B. F. Skinner's Verbal Behavior. *Language*, v.35, n.1, p.26-58, 1959.
- CHURCHLAND, P. Matéria e consciência (Maria Clara Cescato, trad.). São Paulo: Editora Unesp, (Trabalho original publicado em 1984), 2004.
- COELHO, B. *A hipótese da mente estendida: um debate acerca dos limites da cognição*. (Dissertação - Unesp) Marília, 2017.
- COELHO, J. Mente consciente, cérebro e corpo: casos de gêmeos siameses. *PRINCIPIOS*, v. 27, p. 91-113, 2020.
- COLTHEART, M. *The cognitive level of explanation*. Australian Journal of Psychology, v. 64, n. 1, p. 11-18, 2012.

- DAMÁSIO, A. *O erro de Descartes: emoção, razão e o cérebro humano*. Editora Companhia das Letras, 1996.
- DAVENPORT, David. *Computationalism: Still the only game in town*. *Minds and machines*, v. 22, n.3, p.183-190, 2012.
- DE MENEZES, J. Sobre a viabilidade da noção de conteúdo mental estreito. *Princípios: Revista de Filosofia* (UFRN), v.27, n.53, p. 115-134, 2020.
- DENNETT, D. *Consciousness explained*. New York: Black Bay Books, 1991a.
- DENNETT, D. *Ways of establishing harmony*. In Brian P. McLaughlin (ed.), *Dretske and His Critics*. Blackwell, 1991b.
- DESCARTES, R. *Meditações sobre Filosofia Primeira* (F. Castilho, trad.). Campinas: Editora Unicamp, (Trabalho original publicado em 1641), 2004.
- DI PAOLO, E.; THOMPSON, E. *The enactive approach*. *The Routledge handbook of embodied cognition*, p. 68-78, 2014.
- DREYFUS, Hubert L. *O que os computadores não podem fazer: crítica da razão artificial*. A Casa do Livro Eldorado SA, (Trabalho original publicado em 1972), 1975.
- EDELMAN, D., BAARS, B.; SETH, A. Identifying hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition*, v.14, p.169–187, 2005.
- EDELMAN, D.; SETH, A. Animal consciousness: a synthetic approach. *Trends in Neurosciences*, v.32, n.9, 2009.
- FODOR, J. *Psychological Explanation: an introduction to the philosophy of psychology*. Nova Iorque: Random House, 1968.
- FODOR, J. *The Language of Thought*. New York: Crowell, 1975.
- FODOR, F. A. O problema mente-corpo. TCFC3 – Filosofia das Ciências Naturais, São Paulo: FFLCH-USP, (Trabalho original publicado em 1981) 2011. Disponível em: <<http://opessoa.fflch.usp.br/sites/opessoa.fflch.usp.br/files/Fodor-Port-5.pdf>>. Acesso em: 16 de Julho de 2020.
- FODOR, J. Fodor's guide to mental representation: The intelligent auntie's vademecum. *Mind*, v.94, n.373, p. 76-100, 1985.
- FODOR, J. Making mind matter more. *Philosophical Topics*, v.17, n.1, p.59-79, 1989.
- FODOR, J. *The Mind Doesn't Work That Way*. Cambridge: MIT Press, 2001.
- FODOR, J; PYLYSHYN, Z. *Connectionism and Cognitive Architecture: A Critical*

- Analysis. *Cognition*, v.28, 1988.
- FODOR, J; PYLYSHYN, Z. *Minds without meanings: An essay on the content of concepts*. Mit Press, 2015.
- FRANKISH, K. *Illusionism: as a Theory of Consciousness*. Andrews UK Limited, 2017.
- GAZZANIGA, Michael S.; IVRY, Richard B.; MANGUN, George Ronald. *Neurociência cognitiva: a biologia da mente*. Artmed, 2006.
- HAGGARD, P. Conscious intention and motor cognition. *Trends in cognitive sciences*, v.9, n.6, p.290-295, 2005.
- KIM, J. *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT press, 1998.
- LEAL-TOLEDO, G. Ensaio de um fisicista estrito. In: LEAL-TOLEDO, G.; GOUVEA, R.; ALVES, M. *Debates Contemporâneos em Filosofia da Mente*. CEP, v.5852, p.95-112, 2018.
- LECLERC, A. *Uma introdução à filosofia da mente*. Appris Editora e Livraria Eireli-ME, 2018.
- LOW, P.; PANKSEPP, J.; REISS, D.; EDELMAN, D.; VAN SWINDEREN, B.; KOCK, C. *The Cambridge Declaration on Consciousness*. Cambridge, Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals, 2012.
- LEWIS, D. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, v.50, n.3, p.249-258, 1972.
- LYCAN, W.G. *Mind and Cognition: A Reader*. Oxford: Blackwell Publishers, 1990.
- MASLIN, K. T. *Introdução à filosofia da mente* (FJR da Rocha, trad.). Porto Alegre: Artmed, (Trabalho original publicado em 2007), 2009.
- MCCARTHY, J.; HAYES, P. Some philosophical problems from the standpoint of artificial intelligence. In: B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4*. Edinburgh, Scotland: Edinburg University Press, 1969.
- MELO, L.; PESSOA, O. O dualismo interacionista não precisa violar leis de conservação da física. In: MACHADO, Nivaldo; SOBRINHO, Elizeu de Oliveira Santos. *Filosofia da mente, ciência cognitiva e o pós-humano: para onde vamos?*, de Monica Aiub, Maria Eunice Quilici Gonzalez e Mariana Cláudia Broens (Organizadoras). Griot: Revista de Filosofia, v. 13, n. 1, p. 119-121, 2016.
- MENDONÇA, W. *Fisicismo Não-Reducionista: Uma atitude sem conteúdo*

- cognitivo? Sobre o desafio de Bas Van Fraassen. *Principia: an international journal of epistemology*, v.11, n.2, p. 171-186, 2007.
- METZINGER, T. *Being no one: The self-model theory of subjectivity*. MIT Press, 2004.
- NAGEL, Thomas. Como é ser um morcego? *Revista da Abordagem Gestáltica: Phenomenological Studies*, v.19, n.1, p. 109-115, (Trabalho original publicado em 1974), 2013.
- NEWELL, A. Physical symbol systems. *Cognitive Science* 4, 1980.
- NEWELL, A.; SIMON, H. Computer Science as Empirical Inquiry: Symbols and Search. *The ACM Communications*, v.19, n.3, 1976.
- NOË, Alva. *Action in perception*. MIT press, 2004.
- PENNYCOOK, G. Analytic cognitive style predicts religious and paranormal belief. *Cognition*, v.123, n.3, p.335-346, 2012.
- PERINI-SANTOS, E. Contextualismo. In: BRANQUINHO, João; SANTOS, Ricardo. *Compêndio em linha de problemas de filosofia analítica*, Lisboa: Centro de Filosofia da Universidade de Lisboa, 2014.
- PEREIRA JR., A. O Conceito de Sentimento no Monismo de Triplo Aspecto. *Kínesis, Edição Especial – Debate*, v.7, n.15, p.1-24, 2015.
- PLACE, Ullin T. Is consciousness a brain process?. In: *The mind-brain identity theory*. Palgrave, London, 1970. p. 42-51.
- PYLYSHYN, Z. (ed.) *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. New Jersey: Ablex Publishing Corporation, 1987.
- REVONSUO, A. Consciousness, dreams and virtual realities. *Philosophical Psychology*.v.8, n.1, p.35-58, 1995.
- RYLE, G. *O conceito de espírito* (ML Nunes, Trad.). Lisboa: Moraes, 1949/1970.
- RUSSELL, B. *A análise da mente*. Zahar Editores, (Trabalho original publicado em 1921), 1974.
- SACKS, O. *O homem que confundiu sua mulher com um chapéu: e outras histórias clínicas*. Editora Companhia das Letras, (Trabalho original publicado em 1985), 2016.
- SANT'ANNA, A. Funcionalismo homuncular e o problema dos qualia. *Filogênese*. Marília, v.5, p.1-15, 2012.
- SEARLE, J. *Speech acts: An essay in the philosophy of language*. Cambridge university press, 1969.
- SEARLE, J. *Mentes, cérebros e programas*. In: *Cérebros, Máquinas*

- eConsciência. Org. João de Fernandes Teixeira, São Carlos: EDFUSCar, (Trabalho original publicado em 1980), 1996.
- SEARLE, J. *O mistério da consciência e discussões com Daniel C. Dennett e David J. Chalmers*. Paz e Terra, (Trabalho original publicado em 1990), 1998.
- SMART, J. Sensations and brain processes. *The Philosophical Review*, v.68, n.2, p.141-156, 1959.
- SKINNER, B. *Verbal behavior*. New York: Appleton Century-Crofts, 1957.
- TEIXEIRA, J. O que é filosofia da mente. 2ed, Porto Alegre: Editora Fi, 2016.
- TURING, A. *Computação e inteligência*. (FC Hanssem, Trad.) Em J. de F. Teixeira (Org.), *Cérebros Máquinas e Consciência: Uma Introdução à Ciência da Mente*. São Carlos: Editora da UFSCar, (Trabalho original publicado em 1950), 1996.
- TURING, A. On computable numbers with an application to the entscheidungs problem. *Proceedings of the London Mathematical Society*2, v.42, n.1, 1936.
- VARGAS, G. *A consciência na filosofia contemporânea da mente*. 1ª. ed. Novas Edições Acadêmicas, 2014. 164p.
- VELMANS, M. Reflexive monism. *Journal of Consciousness Studies*, v.15, n.2, p.5-50, 2008.
- VITTI-RODRIGUES, M.; MATULOVIC, M.; GONZALEZ, M. Informação-processo e abdução. In: ALVES, Marcos Antonio; MARTÍNEZ-ÁVILA, Daniel; GRACIO, Maria Cláudia Cabrini. *Informação, conhecimento e modelos*, p.131-151, 2017.
- VON NEUMANN, J. The First Draft Report on the EDVAC. *IEEE Annals of the History of Computing*, v.15, n.4, (Trabalho original publicado em 1945), 1999.
- WATSON, J. A psicologia como o behaviorista a vê. *Temas em psicologia*, v.16, n.2, p. 289-301, (Trabalho original publicado em 1913), 2008.

11

Filosofia Política¹²⁷

Everton M. P. Maciel

Nosso objetivo neste capítulo é compreender a filosofia política como um problema prático e não meramente teórico. Vamos percorrer tópicos centrais tentando observar a postura de determinadas correntes filosóficas em torno de assuntos que permeiam a história do pensamento político da antiguidade aos contemporâneos.

1. Qual a melhor das formas de governo?

A diversidade nas formas de governo remonta a um dos mais antigos problemas da filosofia política. Por vezes, o governo é confundido com o Estado em si mesmo, mas a modernidade deve à sistematicidade da obra de Jean Bodin, *De la Republique* (1576), a compreensão de que o Estado representa a soberania, enquanto a forma de ordenamento estatal é, de fato, muito diversificada. Costumamos, a partir disso, admitir que existem diferentes tipos de governos e não formas diferentes de Estado. Uma forma de governo se refere sempre ao conjunto de meios pelos quais o poder soberano é exercido. O governo dos “muitos”, “poucos” ou de “um”, respectivamente, democracia, aristocracia ou monarquia, é o que torna a discussão célebre desde a *História* de Heródoto. Em torno desse assunto, debruçam-se filósofos e pensadores políticos, indo dos extremos de justificar a tirania de um déspota esclarecido às tentativas de saber se a democracia direta tem vantagens em detrimento de sua irmã, a representativa.

¹²⁷O presente texto foi elaborado tendo em mente uma metodologia do tipo *learning based on problems* para a filosofia. Sem as sugestões valiosas do colega Tiaraju Molina Andrezza, padeceria em muitos outros problemas.

1.1. Platão

O período de síntese da filosofia antiga é marcado por Platão e Aristóteles, ambos preocupados com o tema central em torno de qual seria a melhor forma de governo a ser empregada na cidade-Estado. Em Platão (428-347 a.C.), a célebre obra *A República* tem dois livros voltados ao tema, o oitavo e o novo. Uma das mais amplas características de Platão é ser idealista; portanto, não se trata apenas de saber quais, entre as formas de governo conhecidas, é a mais adequada à cidade-Estado grega, mas qual delas é a perfeita. É preciso levar em conta que a perfeição, nesse sentido, requer uma unidade. Em virtude dessa característica temos, na república ideal, um imperativo perfeito do funcionamento do Estado, a tal ponto de que Platão não imagine que possa haver uma corrupção que leve ao desvirtuamento desta forma ideal de governo. Quando Platão, no livro oitavo, faz referência às formas corrompidas de constituição da república, ele trata da sucessão entre elas, todas imperfeitas. A constituição boa, no entanto, não entra nessa sucessão.

Timocracia, oligarquia, democracia e tirania são enumeradas de maneira decrescente, entre as constituições corrompidas. Imediatamente, é preciso notar a ausência de monarquia e aristocracia, pois são seis as formas de governo para Platão. Essas formas de constituição da cidade-Estado não são citadas, pois se encontram entre as formas que podem ser classificadas como perfeitas. Parte da genialidade que Platão fornece de herança para a história da filosofia está justamente na sua maestria de buscar definir conceitos pelo seu contrário. Para não ofender os Deuses, ele não transforma em texto os assuntos da perfeição em si, dedicados à oralidade para sua escola de pensamento. Nesse sentido, Oligarquia é a forma corrompida de aristocracia e tirania a forma corrompida de monarquia. Quando ele questiona: qual é o tipo mais distante da justiça, que podemos opor ao que é mais justo? (545a), tentando identificar os tipos de homens envolvidos em cada uma das formas de governo, sua tentativa é justamente tirar disso o substrato daquilo que ficou conhecido como “rei-filósofo”, o governante da constituição política ideal.

É preciso fazer notar que, para Platão, parece ser indiferente se o governante é um só (monarquia) ou composto por um conjunto de homens (aristocracia). Timocracia, o governo exercido pela força da honra, é tido como um governo de transição entre as formas tradicionais e a constituição perfeita.

A *República* é a descrição da constituição perfeita, permeada pelo método dialético e fiel ao projeto platônico de evitar, a todo custo, cometer heresias, tratando da perfeição inerente ao mundo das Ideias através da palavra escrita. O *Político*, por outro lado, será justamente a busca pela descrição do “rei-filósofo”, o governante perfeito, aquele que possui a ciência do governo justo. O trabalho da filosofia política, desde então, é encontrar o governo que materialize a justiça na comunidade política.

1.2. Aristóteles

A obra *Política* de Aristóteles (384-322 a.C.) usa o termo *politeia* para identificar aquilo que estamos chamando de forma de governo. Neste sentido, a *politeia* é a própria constituição que dá ordem à cidade-Estado, tanto no que se refere às funções do governo quanto à autoridade soberana de quem governa. A sétima parte do seu livro é aquela que merece a maior atenção do filósofo estagirita, justamente onde ele discute as muitas formas de democracia e oligarquia. Nos livros VII e VIII, Aristóteles trata de qual é propriamente a melhor forma de governo. Diferente de Plantão, alvo de suas críticas na segunda parte do livro, Aristóteles vê processos de alternância nas constituições e as observa sendo aplicadas. Para ele, o mundo político ideal não pode entrar em cena, mas está em jogo a ideia de estabilidade das constituições, pois as constituições se alternam entre boas e más: a tirania é o governo monárquico exercido em favor do monarca, a oligarquia visa ao interesse dos ricos, a democracia ao dos pobres; mas nenhuma destas formas visa à utilidade comum (1279a-b).

A alternância sugerida por Aristóteles contrapõe formas boas de governo com seus respectivos desvios. Nesta ordem, temos contrapostas as classificações deturpadas: democracia, oligarquia e tirania como variações, respectivamente, das boas formas de governo: *politia* (variação substantiva do termo *politeia*), aristocracia e monarquia. *Politia* e democracia são desvios da mesma constituição e formas de governo intermediárias, a ponto de Aristóteles afirmar que democracia é um desvio menos mau, porque, de fato, desvia-se pouco da correspondente forma de governo (1160b). O trabalho de Aristóteles tem em vista uma concepção de “viver bem”, com muitas formas internas a cada classificação. Ao tratar de pessoas diferentes em meio aos governados – ricos e pobres, por exemplo – ele estabelece uma distinção entre o público e o privado,

resguardando o espaço do público para a política. A justa-medida dos assuntos éticos não é dissociável, neste sentido, dos assuntos da política, pois a distinção entre ricos e pobres não é meramente numérica, mas substancial. Segundo Aristóteles, tem-se a democracia se os livres e os pobres, sendo em número predominante, são senhores do poder, oligarquia se o são os ricos e os mais nobres, que constituem minoria (1290b).

Para Bobbio, resolve-se esse problema de saber qual a melhor forma de governo observando que *politia* aristotélica é o resultado de uma mistura entre oligarquia e democracia, de duas formas deturpadas de governo: uma na qual aqueles que governam são em sua maioria ricos e bem nascidos e outra na qual aqueles que são pobres. Desse modo, *política* não é meramente o governo do povo ou da democracia em sua forma acertada e correta, mas uma mistura de oligarquia e democracia. Na prática, isso significa que o bom governo de muitos – o terceiro lugar no esquema geral – é um espaço a ser ocupado, uma ideia abstrata a que não corresponde historicamente com nenhum regime já existente (2017, p.45). O problema da interpretação de Bobbio é que tudo vira democracia, mas ele acerta quando sugere que a justa-medida aristotélica é parte intermediária do esquema geral, pois parece ser importante, em meio à alternância das constituições entre boas e más, admitir a centralidade da “menos má”.

Os manuais de filosofia convencionaram o tratamento de Aristóteles como um aristocrata, em virtude de uma confusão intermediária entre sua posição política, que permaneceu e permanente em diálogo com o meio-termo, defendido por ele na ética. Independente disso, é impreterível levar em conta que o projeto de ter uma política estável não era depositado nos extremos do esquema das diferentes formas de constituições. Diferentemente de Platão, ele não procurava a constituição perfeita, mas a mais adequada em meio às conhecidas.

1.3. Maquiavel

No seio do renascimento, Nicolau Maquiavel (1469-1527) foi o fundador do realismo político, uma forma de lidar com os assuntos da natureza do Estado, voltada para decisões pragmáticas, em torno da sua formação e manutenção. Se, entre os gregos, havia a perspectiva de observar as diferentes constituições das

idades-Estado, Maquiavel parte de pressupostos retomados a partir da *res pública* romana. Essa postura é nova, pois o período de síntese da história da filosofia nos ensinou a analisar as constituições dos diferentes tipos de governo, de acordo com as variações numéricas em torno de quantas pessoas governavam e qual é a natureza do governo analisado (bom ou mau, justo ou injusto). Na sua obra capital, *O Príncipe*, Maquiavel sentencia, nas duas primeiras linhas, a diferença fundamental entre o seu trabalho e o dos seus antecessores: “todos os estados, todos os domínios que tiveram e têm império sobre os homens foram e são ou repúblicas, ou principados” (2010, p.5). Essa diferença é fundamental porque estabelece que a administração dos assuntos de Estado é responsabilidade, apenas, de um ou mais de um. Enquanto, para os antecedentes, a especulação em torno dos Estados existentes possuía uma variação de saber se o governante ou o conjunto de governantes era bom ou mau, análise indispensável para reconhecer se estamos diante de uma monarquia ou de uma tirania, Maquiavel as resume em principados, tomados como o reinos, ou repúblicas, aqui pensadas no sentido de aristocracia/oligarquia ou de democracia.

Não é incomum que os intérpretes de Maquiavel distingam o autor de *O Príncipe* de um Maquiavel “republicano”, como se isso representasse a diferença entre alguém que estabeleceu um manual com recomendações autoritárias e impiedosas, embora possa também ser considerado, em sentido oposto, um democrata. Em parte, essa aparente incongruência acontece pela distinção entre ele e os autores precedentes a respeito da variedade das formas de governo, mas também porque sua obra mais famosa possui um caráter bastante enigmático, falando das qualidades do bom governante e das condições nas quais a política opera na prática, *virtù* e *fortuna*. *Virtù* diz respeito às qualidades pessoais de dominar eventos e realizar, através de qualquer meio, as finalidades para a manutenção do Estado. *Fortuna* é a compreensão do conjunto de eventos que independem dos desejos humanos, inevitáveis por natureza, mas que precisam ser compreendidos pelo bom governante. O sucesso em alcançar os fins é o critério fundamental para se distinguir a boa da má política, e disso advém o princípio maquiavelano¹²⁸ de que os fins justificam os meios.

¹²⁸ Usamos “maquiavelano” ao invés de “maquiavélico”, para preservarmos o uso teórico como separado do uso popular do termo.

O realismo político depende, em larga medida, do pano de fundo de que existe uma constância nas características da natureza humana, uma ordem estabelecida previamente, que deve ser compreendida e utilizada na boa forma de governo. Quando Maquiavel é apontado como o fundador da ciência política, se trata de alguém que examina essa ordem que compõe acontecimentos do passado e torna-se capaz de tomar boas decisões, como se pudesse prever o futuro dos acontecimentos, levando em consideração a forma como os súditos ou cidadãos reagem, a maneira como os apoiadores e opositores do governo se interessarão ou se afastarão dos negócios públicos. Esse Maquiavel “republicanizado” é presente na obra *Discurso: comentários sobre a primeira década*, de Tito Lívio:

Quem estudar a história contemporânea e da antiguidade verá que os mesmos desejos e as mesmas paixões reinaram e reinam ainda em todos os governos, em todos os povos. Por isso é fácil, para quem estuda com profundidade os acontecimentos pretéritos, prever o que o futuro reserva a cada Estado, propondo os remédios já utilizados pelos antigos ou, caso isso não seja possível, imaginando novos remédios baseados nas semelhanças dos acontecimentos (1994, p.129).

A ciência política, desde então, é essa análise empírica do passado que qualifica inclusive o presente, tendo em vista o sucesso da política empregada. A questão comum a toda obra de Maquiavel é a instabilidade, a discórdia e o conflito. Evitar as “doenças” da política é a tarefa do bom governante. Essa compreensão científica deixa de lado preceitos de justiça e de bem viver para analisar o sistema político como um corpo que deve ser compreendido a partir dos seus problemas físicos, tal qual um médico que estuda os males do corpo para propor o remédio.

Os estudos de Maquiavel advindos dos romanos são tão negativos e preocupados com os males do governo quanto aqueles da sua obra mais famosa. Nesse sentido, o Maquiavel do *Discurso* não é diferente daquele que escreveu *O Príncipe*, mas os objetos do estudo são diferentes. A melhor forma de

constituição do governo, em ambos os casos, será sempre aquela que trabalha com o que há de melhor nas experiências dos principados e das repúblicas, uma forma mista entre as duas constituições possíveis. As palavras da sua análise sobre a debandada dos romanos são esclarecedoras:

Unido, o povo é forte; isolados, os indivíduos são fracos [...] De fato, quando os ânimos se arrefecem, e todos percebem que é preciso voltar para casa. Esvai-se a confiança que a massa depositava na sua força; cada um pensa na própria salvação, decidindo-se a fugir, ou a trair. O povo que deseja evitar tais perigos deve escolher um chefe que o dirija, o mantenha unido, e o defenda. Foi o que fez a plebe quando abandonou Roma depois da morte de Virgínio, nomeando vinte tribunos para zelar pela sua segurança (1994, 178).

É difícil considerar nessas observações que exista um antagonismo entre a nobreza e os pobres; Maquiavel antecipa em muitos séculos uma concepção de sociedade civil, comprometida com “a coisa pública” para a manutenção de suas liberdades básicas.

1.4. Montesquieu

A forma constitucional com que Charles-Louis de Secondat, Barão de La Brède e de Montesquieu (1689-1755) tratou da monarquia deixou como legado o sistema de tripartição de poderes mais influente da história dos governos modernos. Em grande medida, a tripartição dos poderes em Executivo, Legislativo e Judiciário se deve a compreensão que Montesquieu empregou das alterações provocadas no seio da sociedade francesa entre os séculos XVII e XVIII, que culminariam na Revolução.

Montesquieu apostou que, dentre todas as constituições disponíveis, a excelente deveria ser a monarquia constitucional, com os poderes do monarca sendo moderados por um complexo sistema de freios, pesos e contrapesos, advindo da sociedade civil. Cabe ressaltar aqui que a compreensão dele em

torno da ideia daqueles que compõem a sociedade civil é diferente da apresentada mais tarde por Hegel como “burguesa”, com um sistema de privilégios focado nas vantagens financeiras. No *Espírito das Leis*, vemos, sim, uma sociedade civil apartada do Estado, mas com o foco em “leis fundamentais” que busquem unificar as diferenças, realizando o interesse comum, a partir de uma referência social. Em larga medida, isso incluía uma sociedade civil unificada na administração pública dos poderes Legislativo e Judiciário, não apartada do Estado, no sentido hegeliano.

As concepções de honra e desigualdade eram os temas emergentes para se definir uma boa forma de governo. O primeiro em virtude de sua subjetiva força de unidade em torno da existência do próprio Estado; o segundo conceito, desigualdade, como tratamento para corrigir falhas e privilégios internos às cortes, aspecto objetivo de maximização da desigualdade e impeditivo da modernização rumo a uma monarquia constitucional. A sociedade civil, nesse aspecto, é aquela em que cada concidadão pode e deve aproveitar sua singularidade em relação aos outros, diferente de uma perseguição egoísta dos próprios fins individuais. Nesse aspecto, sua perspectiva de um bom governo envolve certa complexidade e pluralidade na confecção das leis. Tanto a liberdade dos cidadãos é atrelada à ideia de fazer o que as leis permitem, no sentido de uma garantia positiva da vida civil equilibrada, quanto o governo precisa de um sistema de freios em busca do equilíbrio entre poderes. A famosa tripartição dos poderes em Montesquieu busca justamente que um poder não se torne mais forte que outro, especialmente aquele ocupado pelo monarca, com o esvaziamento de diversas funções de Estado. Todos os poderes, na sugestão de Montesquieu, veem suas prerrogativas para o funcionamento da máquina pública limitados. Abre-se espaço para uma disputa em torno dos privilégios da administração pública que não são honrosos, demanda fundamental em torno da união que mantém o governo estável.

Montesquieu tornou a tripartição das formas de governo algo desinteressante para a filosofia política. Ele foi o primeiro a compreender que o importante é a garantia da liberdade dos cidadãos dentro do Estado, pela limitação dos poderes do próprio Estado, por meio de uma redistribuição equânime dessa atividade de governar, uma tripartição interna ao próprio governo. O bem-estar dos cidadãos independe da forma de governo, mas está ligado à participação do cidadão na forma de governar.

2. Relação entre religião e Estado: o poder sacerdotal é equiparável ao poder estatal?

É um equívoco tratar a história da filosofia como tendo uma brecha de mil anos, desde a queda de Roma até o ressurgimento das letras no Renascimento. Ignorar o atraso provocado pela influência da Igreja no desenvolvimento da política real também parece ser fechar os olhos para o óbvio. O mais cético de todos os filósofos, David Hume, fazendo o papel de historiador, reconhece o seu papel como “outra trincheira contra a autoridade real, ao mesmo tempo em que provocava muitas inconveniências e danos. O excelso clero talvez não fosse tão propenso à violência quanto os barões, mas como aspirasse à completa independência em relação ao Estado e tivesse a religião para acobertá-lo, mostrou-se um obstáculo à constituição do reino e à execução das leis” (2017, p.113). Esse panorama de Hume coloca a Igreja da Idade Média como um tipo de poder moderador, que não possuía nenhuma outra referência institucionalizada, antes da ascensão do cristianismo como religião predominante na Europa. O poder da Igreja, na prática de relações, foi o único historicamente comparável ao poder do próprio Estado.

Não devemos esquecer que o desenvolvimento funcional do Estado se deve especialmente ao cristianismo como um componente do poder político medieval. A concepção de lealdade – do vassalo para com o senhor feudal, da nobreza para com o soberano – tem um componente cristão forte. Quando Carlos Magno foi coroado imperador do Sacro Império Romano, em 800, pelo papa Leão III, o poder da Igreja cresceu em todo o ocidente e tornou-se imperativo o reconhecimento da realeza pelo poder papal. A proteção da Igreja de Roma foi exercida a partir da constituição dos impérios e do préstimo de seus serviços militares. A consolidação dos reinos, por outro lado, dependia muitas vezes do reconhecimento de Roma.

2.1. Santo Agostinho

O nome mais influente da chamada filosofia patrística foi Agostinho de Hipona (354-430 d.C.). Convertido, ele foi responsável pela obra *Cidade de Deus*, sua perspectiva ideal de comunidade cristã. “Trata-se do Estado, ou da

comunidade política que é a maior e abarca todas as outras, aponta para o bem em um grau mais elevado que qualquer outra, e para o maior bem” (AGOSTINHO, *apud*, LADNER, 1947, p.403). Não basta apontar a direção da comunidade política com o “maior bem”, como faria uma teoria aristotélica, por exemplo. Para Agostinho, era preciso perseguir esse objetivo de vida em comunidade, não como um mero Estado, instável e transitório, mas como uma sociedade supranatural na própria terra.

O maior de todos os bens, Deus, ocupou espaço entre os mundanos, através da revelação de Cristo. Agostinho tinha isso em mente quando sugeriu que perseguir o maior de todos os bens seria o mesmo que formar uma comunidade mundial cujo seus membros pudessem abraçar o reino dos céus.

A Igreja erigida com base nessa revelação de Deus, que viveu entre os mortais, para dar a si mesmo em remissão de nossos pecados, é a “única sociedade humana engajada em construir a Cidade de Deus” (*idem*, 404). A concepção de Estado de Agostinho é indiferente às comunidades territoriais. Mesmo o Império Romano cristianizado é envolto em valores demasiadamente relativos. A *res publica* de Cícero, por exemplo, sem a verdade da justiça através de Deus e dos homens, não pode ser considerada uma república no sentido amplo que essa constituição adota, sugerido por Agostinho. O mesmo se passa com a *polis* de Aristóteles, que, para ele, não possuía uma referência sólida para se tornar o maior dos bens políticos. Ele então acaba por negar qualquer comunidade, exceto a Cidade ou Sociedade de Deus, a personificação do bem público.

Agostinho foi o grande mentor da institucionalização da Igreja e, conseqüentemente, do catolicismo como poder paralelo em meio aos estados europeus. A ascensão do clero, o fortalecimento dos bispados e o estabelecimento da consolidação de uma hierarquia rígida para a consolidação do poder da Igreja de Roma deveu-se em larga medida a compreensão de sua obra. Não significa, no entanto, que ele tenha ignorado conceitos antigos da filosofia política grega e romana. Ele aceita certos elementos valiosos para a concepção de Estado dos antigos. Lei (direito) e justiça, ordem e paz são alguns exemplos importantes que percorrem toda a obra do Santo. Conceitos de verdades metafísicas destinados à vida pública das pessoas são importantes para a doutrina cristã. A cristianização abraçou essas ideias pagãs, tentando oferecer uma verdade integral a respeito do seu conteúdo. Reino e Igreja,

mesmo ainda distantes da realização completa dos seus ideais, não devem ser considerados meramente como comunidades ou territórios, mas formas e designios de um governo justo. Essa condição mista da Cidade de Deus é a que temos disponível na terra.

2.2. São Tomás de Aquino

São Tomás de Aquino (1227-1274) foi o grande responsável pela “cristianização” da obra de Aristóteles na idade média. Seu papel foi trabalhar com a organização das verdades da fé cristã e, por assim dizer, harmonizá-las com a síntese aristotélica da filosofia, inclusive a filosofia política. Para ele, não deve haver ponto de conflito entre fé e razão. Assim, os poderes régio e papal devem ser justificados, ao mesmo tempo, tanto pelo conhecimento universal – o conjunto de todas as verdades da fé cristã –, quanto pelo conhecimento crítico – com as condições do conhecimento humano bem delimitadas, sabendo até que ponto pode chegar a compreensão.

A distinção entre os poderes do Estado e da Igreja é apenas convencional para São Tomás, mas eles não representam antagonismos. O assunto é tratado na obra *De Regno*. Trata-se da diferença entre a compreensão da natureza do direito e da moral. Essas representações da ordem intelectual são institucionalizadas, para prover ao mundo ordem civil e espiritual, respectivamente.

Na sua obra fulcral, *Suma Teológica*, seu compromisso passa por uma organização epistemológica do mundo, pois o conhecimento verdadeiro vem de adequar a inteligência à coisa. Em última análise, o ponto de convergência para uma vida boa, para Tomás, não é encontrado nos bens da alma, tampouco nos bens corpóreos, mas na contemplação da verdade. A felicidade do homem, nesse sentido, depende do esforço que ele faz para compreender as verdades da fé e não apenas aceitá-las como dogmas. Ele mostra que há, em Aristóteles, uma filosofia verdadeiramente autônoma, racional e independente do dogma, mas em harmonia com o motor imóvel, ou causa primeira, reconhecido pela tradição cristã como Deus.

É muito significativo que São Tomás, na sua discussão da relação entre Estado e Igreja, ignore o Sacro Império Romano, com exceção de um precedente: um comentário sobre a segunda carta aos tessalonicenses. Nela,

São Tomás comenta que o Sacro Império Romano-Germânico tem mais referência com a eternidade do poder eclesial na Terra do que a última fase do Império Romano na Antiguidade ou do que um cristianismo original, descontinuado pela queda do Império, na Antiguidade. Ele emprega um senso de continuidade, para demonstrar a importância institucional do poder da Igreja. Uma boa interpretação para isso vem de Ladner: “[...] São Tomás, ou um de seus pupilos claramente estabelece que o Império Romano foi transportado da ordem temporal para a espiritual (*commutatus de temporalis in spirituale*) e infere que a profecia diz respeito a essa eternidade, agora aplicada à Fé Católica da Igreja Romana” (1947, p.420).

O poder temporal, do qual trata o governo do monarca, é dedicado a dirigir a vida virtuosa, materializada pela comunidade política. Isso não significa que o fim do homem se esgota na cidade. É preciso considerar o caminho apresentado pelas verdades da fé, a ortodoxia da Igreja. Para Tomás, é por este caminho “por onde o homem cristão tem precisão daquele cuidado espiritual, pelo qual seja dirigido ao porto da salvação eterna, cuidado esse proporcionado aos fiéis pelos ministros da Igreja de Cristo” (1997, p.163). O meio para alcançar a felicidade eterna é a bem-aventurança da vida celeste. Nesse ponto, o poder da Igreja de Cristo é superior ao poder temporal. Cristo, por meio de Pedro, instaura a supremacia do poder papal e, diferentemente das doutrinas pagãs, mas em consonância com elas, a nova lei torna os reis sujeitos aos sacerdotes.

2.3. João Quidort de Paris

É provável que a primeira sugestão de secularização ou do afastamento entre o poder do eclesiástico e estatal tenha surgido em meio a própria Igreja com um contemporâneo de Tomás de Aquino. João Quidort (1255-1305) provocou um esvaziamento do conceito metafísico de natureza, uma ruptura com o período de síntese da história da filosofia (Platão e Aristóteles), e um esboço de liberalismo político. Os princípios expressos na obra *De Potestate Regali et Papali* são baseados em um conjunto bastante audacioso para a época: a) separação das esferas de poder eclesiástico e político; b) impossibilidade de exercício de poder sobre a propriedade dos bens (cobrança de impostos, confiscos, adjudicação do litígio) por parte do clero (PIERPAULI, p.125).

A teoria de João se sustenta em uma máxima imperativa do exercício

do poder, sustentada no direito de propriedade e aquisição dos bens. Diferente de Tomás, ele não estabelece uma justificação interna ao próprio sistema, mas recorre a um elemento externo, disposto na natureza das coisas, para fundar sua teoria política. Ele observa o funcionamento regular das relações de posse, em que o dono pode simplesmente dispor da coisa possuída de acordo com suas próprias demandas, sem obrigação de ceder para o uso comum. A esse respeito, João faz sua própria interpretação do livro I da *Política* de Aristóteles. O erro dos teólogos anteriores está em admitir que Cristo seria um rei na terra, com interferência imperativa no direito de propriedade. “Dai, pois, a César o que é de César, e a Deus o que é de Deus” (MATEUS, 22:21), para João, é uma demonstração de que os assuntos de propriedade terrena não devem ser regulados pela Igreja. Se convencionou que o papado é uma extensão dos direitos de Cristo, no que se refere aos seus domínios, com base em uma ideia equivocada da jurisdição desses domínios. Assim, a Igreja de Roma teria poder de interferência sobre os bens temporais dos reinados e cortes. João foi o primeiro a contestar essa jurisdição da Igreja.

Enquanto a tradição semeou a ideia de que o Papa tem um poder oriundo diretamente de Deus e os príncipes têm um poder mediado pelo Papa, João define ambos em esferas separadas diametralmente uma da outra. O reino é o governo de uma massa organizada tendo em vista o bem comum. O sacerdócio, no entanto, é o poder espiritual. O clero é responsável pela administração dos sacramentos aos fiéis. Essa organização eclesial tem em vista apenas um chefe superior. Trata-se de um dispositivo hierárquico mais adequado aos ministros da Igreja do que aos príncipes, porque o clero é ordenado especialmente para o exercício do culto divino, pertence a Deus como povo escolhido. Os leigos fiéis, nesse mesmo aspecto, não recebem uma determinação direta que os obrigue, diante das coisas temporais, a subordinar-se a um só monarca (PIERPAULI, p.126s).

O instinto natural dos cristãos não ordenados acaba por orientá-los a viver em comunidades políticas para alcançar o bem comum. A divisão entre o poder sacerdotal e estatal sugerida por João leva em conta que a comunidade de cristãos elege ou aceita chefes para alcançar seus objetivos. Não é possível tratar, portanto, a comunidade de cristão como se fosse a comunidade sacerdotal, diante de um único soberano. No âmbito dos governos, João visualizava uma política muito mais multifacetada que exigia um ordenamento natural, diferente

da Igreja.

3. Contratualismo: o contrato social é real ou hipotético?

Uma das mais influentes teorias moral e política da história da filosofia é o contratualismo. Trata-se do reconhecimento da origem do Estado ou da comunidade civil a partir de uma convenção, ou contrato entre seus membros. O contratualismo é empregado tanto para justificar o poder absoluto, quanto para traçar os limites do poder político legitimamente exercido. Uma das polêmicas que percorre a corrente desde o seu surgimento entre os antigos é saber se Licofronte, sofista e discípulo de Górgias, estava levantando uma tese hipotética ou acreditava que os seres humanos, de fato e historicamente, constituíram assembleia, ao afirmar que a lei é puramente uma convenção para a “garantia dos direitos dos homens” (ARISTÓTELES, 1280b).

Eclipsado na Idade Média, por conta das teorias divinas da origem do Estado, o assunto foi retomado apenas no final do renascimento, no início da modernidade. Em virtude disso, vamos procurar indício de um contrato social real ou hipotético em Hobbes, Locke e Rousseau, os principais defensores modernos dessa corrente. Muitos dos contratualistas contemporâneos, incluindo Rawls, seguem a linha de que o contrato é uma construção teórica que busca uma justificação, não limitada historicamente, para o estabelecimento das regras nas sociedades políticas. Observar como isso se deu nos três principais nomes do contratualismo normativo pode contribuir para compreensão geral da justificação do Estado.

3.1. Thomas Hobbes de Malmesbury

Thomas Hobbes de Malmesbury(1588-1679) foi o filósofo inglês que deu ao contratualismo a dimensão ampla que ele tem desde então, enquanto filosofia política e moral. Ele defendeu um modelo em que o Soberano ocupa um espaço de um poder absoluto, frente aos contratantes, o povo. Sua perspectiva do contrato envolve uma dimensão irreversível na qual todos transmitem o direito de governar-se ao Soberano. A alegoria que Hobbes resgatou para essa entidade estatal absoluta é o Leviatã. Cabe ressaltar, de pronto, que é indiferente para a teoria de Hobbes se estamos nos referindo a um indivíduo, monarca, ou a

uma assembleia de homens. O contratualismo, nesse aspecto, fornece ao Leviatã as características de um Deus mortal com autoridade para prover a paz e a defesa entre os homens. Essa autoridade conferida pelo conjunto de indivíduos que compõem o Estado é absolutamente imperativa. Tem tanta força e poder que pode inclusive disciplinar a vontade de todos para a conquista da paz.

Por natureza, todos os homens são iguais. Para Hobbes, as diferenças físicas não são significativas e podem ser anuladas e superadas pelo emprego da astúcia ou coleta de apoio entre seus pares. Ele acredita em uma uniformidade, entre os homens, das qualidades e das vontades de alcançar objetivos. No Estado de Natureza, em independência de cada homem com relação a todos os outros, a guerra de todos contra todos é a máxima de conduta das ações humanas. Não existe justiça ou injustiça em uma condição tão miserável e desprovida de paz, na qual cada homem pode saquear e pilhar seu vizinho, à revelia de qualquer necessidade de explicação ou justificação. Nesta condição de profunda igualdade e liberdade naturais, nenhum homem está obrigado a cumprir qualquer pacto estabelecido e se o faz é por força do medo e não por respeito ao acordo.

A natureza criou os homens tão iguais em suas faculdades físicas e mentais que, embora alguns homens sejam de fato mais fortes fisicamente, ou mentalmente mais rápidos, ainda assim, ao somarmos tudo, as diferenças entre os homens não são tão grandes a ponto de alguém poder reivindicar somente para si qualquer vantagem que outra pessoa também não a possa reivindicar (HOBBS, 2015, p.115).

[...] então, no estado de mera Natureza (que é o estado de Guerra de todos contra todos), o Pacto será Nulo caso exista qualquer suspeita razoável; Mas se houver um Poder comum que reine sobre ambas as partes, com direito e força suficientes para obrigar o seu cumprimento, o Pacto não será Nulo (*idem*, p.127).

Apenas o pacto efetuado por medo obriga as partes (*idem*, 129). Esse ponto é fundamental para compreender que a condição de funcionamento da

teoria contratualista independe da existência real de um período histórico no qual os homens tenham de fato contratado. Trata-se, outrossim, da condição natural da existência humana que precisa ser compreendida a partir do direito natural e das leis da natureza.

O direito natural (*jusnaturale*) é aquele que confere a qualquer indivíduo o uso de sua liberdade da forma que bem lhe aprover. Esse direito inclui tudo aquilo que estiver à sua disposição para preservar a sua própria natureza. Já as leis naturais (*lexnaturale*) incluem a razão, uma postura que impede cada um de fazer aquilo que lhe seja destrutivo à própria vida ou remova os meios de preservá-la. O ponto fundamental é que, se levarmos esses conceitos importantes do vocabulário hobbesiano em conta, “todo homem tem direito a tudo”. A primeira lei da natureza, guiada pela razão, é a de fazer tudo aquilo que for possível para buscar a paz e defender-se por todos os meios possíveis. Renunciar ao direito de tudo possuir torna-se a segunda lei imperativa da natureza, pois, sem isso, mantém-se o “estado guerra de todos contra todos”. O homem do contrato social é aquele que, fazendo uso da razão, reconhece que direito é algo que possuímos, enquanto lei é algo que obedecemos. Para Hobbes, ninguém dá a outro o direito a qualquer coisa, apenas estamos em condições de renunciar a direitos. É dessa renúncia de uma liberdade irrestrita que nasce o contrato social.

O homem abre mão da sua liberdade natural de tudo fazer, disponível em estado de natureza. Esse excedente da liberdade de cada um dos contratantes é depositado no Leviatã. É importantíssimo salientar que Hobbes não espera que o direito natural desapareça com o estabelecimento do pacto; igualmente, as leis da natureza não perdem seu valor imperativo. A ideia de que podemos fazer qualquer coisa que esteja a nossa disposição para evitar a própria destruição, de que não podemos ser obrigados a deixar de resistir em caso de atentado contra nós mesmos, é fundamental. O pacto social de Hobbes não envolve um “antes ou depois” do contrato, estado de natureza como contraposto ao estado civil, mas é algo permanente como regra tácita no âmago de qualquer constituição formada em torno do governo.

Podemos imaginar que tal período nunca tenha existido, nem estados de guerra como esses; e eu acredito que, em geral, nunca tenha sido assim

mesmo, no entanto, existem lugares em que se vive desse modo hoje em dia; pois os povos selvagens de muitas partes da *América* não têm nenhum governo, exceto o governo de pequenas Famílias, cuja concórdia depende da lascívia natural [...]

Mesmo que nunca tenha existido tal tempo em que os indivíduos estivessem em estado de guerra de todos contra todos, mesmo assim os Reis e as pessoas que possuem Autoridade soberana relacionam-se dessa forma o tempo todo (*idem*, p.199).

Antes de ser um estado de natureza histórico real, Hobbes estabeleceu com isso os fundamentos de uma hipótese de trabalho para a filosofia política.

3.2. John Locke

John Locke (1632-1704) foi uma das principais influências do liberalismo inglês e está entre os intelectuais mais importantes das Revoluções Americana e Francesa. Foi usado de maneira controversa tanto por revolucionários quanto por conservadores. Locke estreou a filosofia política inglesa no terreno do empirismo e foi o grande ideólogo do materialismo, deixando suas pegadas inclusive nos trabalhos mais antagônicos ao seu, como Karl Marx, herdeiro de uma concepção material de propriedade privada. Entre os seus interesses diretos está a autoridade do Estado baseada na religião; a lei natural como responsável por sancionar a autoridade Estatal; e o fundamento da lei natural na experiência do conhecimento humano.

Nos seu *Dois Tratados Sobre o Governo*, Locke compreendia que a natureza da sociedade civil exigia o poder da magistratura, sendo ele necessariamente absoluto, especialmente se levarmos em conta a ideia de que a defesa de propriedade, por natureza, é um bem do qual não podemos abrir mão. Por certo, há um elemento minimalista presente na ideia de propriedade privada, sendo que o corpo é o elemento atômico do qual o homem não pode se desfazer. “Embora a Terra e todas as criaturas inferiores sejam comuns a todos os homens, cada homem tem uma *propriedade* em sua própria pessoa. A esta ninguém tem

direito além dele mesmo” (LOCKE, II, §27).

Ao mesmo tempo em que não duvidava da superioridade dos pais frente aos filhos, velhos frente aos mais jovens e do monarca diante de todos, Locke também via nisso um problema empírico, possível de ser superado com a observação, astúcia e capacidade de melhora do entendimento. A lei da natureza estabelece os limites da liberdade natural. Se não houvesse limites para a lei, também não haveria para a liberdade. Justamente por essa característica, é possível que um filho ultrapasse em superioridade um pai e assim por diante. Longe de advogar um aristocratismo paternalista, ou qualquer outro tipo específico de governo, ele depositou no contratualismo a origem do surgimento da sociedade civil, dependente de um senso naturalizado de propriedade e sucessão hereditária. Mesmo que não tenha descrito um estado de natureza originário, tampouco advogado que tenha existido um “estado de guerra”, antes do surgimento da sociedade civil, Locke tinha na imagem do monarca a subordinação de toda a vontade humana, em virtude da ordem disposta na lei da natureza.

O direito do governante é convertido em uma questão de vontade, uma vez que Locke não vê o gênero humano como diferente dos animais, do ponto de vista material. Em parte, isso se justifica pela questão da vontade, pois todo aquele que almeja o poder e os direitos monárquicos pode ser tratado como um animal. A vontade humana, nesse sentido, está muito mais próxima de um elemento materialista do que de uma ideia de livre-arbítrio e a fonte do poder não pode ser arbitrária ou revelada.

As especulações de que as primeiras formações de governo são provenientes da tentativa de se defender em caso de guerra, estando o general a comando dos homens por sua experiência militar, mesmo que a forma original de famílias fosse reestabelecida tão logo a paz voltasse a reinar, serve de ponto de partida. A formação histórica do primeiro estado, por outro lado, é indiferente para a teoria naturalista que Locke estava estabelecendo, por meio da ideia de igualdade natural da racionalidade, elemento comum a todos os homens que pactuam para se autopreservarem.

Um individualismo radical explica o fato de que existem governos, e a agressão sofrida por um indivíduo é pensada como uma agressão comum a toda a espécie. O direito de defender a propriedade é natural à toda humanidade. Mesmo que esse poder, conferido pela lei da natureza, seja coletivo, ele pode

ser exercido por uma assembleia de homens ou mesmo por um único monarca. Ao mesmo tempo em que o direito de governar é um direito, também, é um poder natural, fundamental e individual, como o direito de preservar a si mesmo e o resto da humanidade.

A natureza nunca permitiria que um homem abandonasse a si mesmo a ponto de descuidar da sua própria preservação. E já que não pode tirar a própria vida, tampouco pode dar a outro o poder de lhe tirar. E que ninguém pense ver nesse direito o fundamento de uma desordem perpétua, pois ele não opera, até que o inconveniente seja tão grande que a maioria o sinta e dele se canse, julgando necessário remediá-lo. Desse perigo, porém, o poder executivo e os príncipes sábios devem guardar-se; e é a coisa, dentre todas as demais, que mais precisam evitar, pois é a mais perigosa (LOCKE, II, §168).

Onde quer que haja autoridade coletiva e permanente, também há o resultado da reflexão dos homens e das medidas por eles deliberadas, um domínio garantido da racionalidade e das leis naturais. O estado de natureza ao qual Locke se refere não é um estado de guerra, como em Hobbes. A guerra é um incidente; não pode definir, portanto, o estado de natureza. Tal qual no direito internacional, mesmo que a condição seja de guerra, o Estado deve ser definido, apenas, como estado de natureza, no qual cada Estado é um indivíduo.

A propriedade é o mais elementar dos conceitos políticos, para Locke, e inédito, até então. O direito de possuir é facultado à espécie humana, como um todo, e não a indivíduos isoladamente. A extensão desse direito configura o Estado. Trata-se, fundamentalmente, da propriedade de sua própria pessoa; trabalho do corpo; obras das mãos etc. Tirar algo do estado de natureza, misturando algum trabalho, gera um tipo de propriedade. Regular e preservar a propriedade faz parte das atividades mais elementares do Estado político, o que inclui a vida humana. Em Locke, a propriedade simboliza os direitos em sua forma concreta, elemento ligado à liberdade, pois estabelece o tema tangível dos poderes e atitudes de um indivíduo. A teoria da propriedade de Locke divide o

mundo em abstrato e concreto. No campo da realidade abstrata, há liberdade e igualdade, entre os homens e Deus; na concretização do ambiente político, há espaço para liberdades políticas e acordos.

O contratualismo de Locke não é apenas uma hipótese de trabalho, porque carrega consigo um elemento materialista concreto, com um surgimento conceitual e material circunscrito ao tempo histórico: a propriedade privada, seu surgimento e manutenção. Por outro lado, Locke apela para uma lei da natureza hobbesiana e reposiciona o conceito de propriedade na própria vida do contratante. Essa característica dá a ideia material de propriedade uma perspectiva abstrata, indivisível, passível de ser operada apenas em uma teoria política com essas características.

3.4. Jean-Jacques Rousseau

Um dos mais importantes nomes do moralismo francês, Jean-Jacques Rousseau (1712-1778) foi responsável por reconciliar as ideias de liberdade e autoridade, além de ter sido uma influência direta para os políticos da Revolução Francesa. O *Contrato Social* é sua obra política mais difundida. Soma-se a ela, textos de política econômica, comentários políticos sobre governos e o célebre *Discurso sobre a Origem e o Fundamento da Desigualdade entre os Homens*. No *Discurso*, escrito para um concurso da academia de Dijon, Rousseau usou toda sua elegância textual para construir a mais canonizada imagem do “bom selvagem” da história da filosofia política. Para ele, o contratualismo de Hobbes parte de um pressuposto não apenas errado de natureza humana, mas contrário ao que de fato pode ser encontrado entre os nativos do Novo Mundo, a materialização concreta da ausência de Estado, e onde o homem, de fato, vive em estado de natureza.

Rousseau utilizou o conceito de “vontade geral” para empregar seus esforços de atingir uma realidade coletiva da qual o contrato social é dependente. A primeira distinção importante que ele estabelece da teoria de base hobbesiana é entre o poder e a vontade. O poder político pode ser transmitido, entregue ou mesmo usurpado; por outro lado, a vontade dos indivíduos serve como mediação para o contrato social, através de uma democracia direta, na qual a vontade de todos se relaciona com o poder político, sem intermediários.

Assim como não partiu de pressupostos envolvendo a liberdade

negativa – de ser permitido fazer qualquer coisa, contanto que esteja ao seu alcance (Hobbes) –, Rousseau também não fundou no direito de propriedade (Locke) a natureza humana: para ele, o primeiro e mais elementar sentimento humano, a existência de si mesmo, é aquilo que define nossa natureza, seguido de outros sentimentos emancipatórios, mas não políticos, no sentido do uso da força. A conservação de si mesmo é o segundo sentimento humano em estado de natureza que ainda é perfeito e intocado pelas preocupações concernentes à vida na sociedade civil. O orgulho ou a inveja, nascentes da comparação do ser humano com seus semelhantes, amor e desejo pelo próprio bem-estar, noção de futuro e provimento de garantias de conforto são outros pontos importantes nesse caminho que a vontade humana ocupa no terreno percorrido em direção ao contrato. Em nenhum momento, por outro lado, Rousseau vê nessas atividades um percalço conturbado, violento, muito menos uma “guerra de todos contra todos”. Pelo contrário, o homem em estado de natureza é sereno, sem preocupações e necessidades para serem atendidas, avesso à vida coletiva e sem nenhuma preocupação que esteja além do presente.

“Essa ideia de propriedade, dependente de muitas ideias anteriores que apenas poderão nascer sucessivamente, não se formou de imediato no espírito humano” (ROUSSEAU, 2000, p.307). O filósofo trata do surgimento da sociedade civil como algo atrelado à propriedade, mas não reduz a natureza humana a um conceito de direito, com o uso tão mais recente entre os homens.

Ao primeiro que tendo cercado um terreno lhe ocorreu dizer ‘isto é meu’ e encontrou pessoas bastante simples para acreditar em si foi o verdadeiro fundador da sociedade civil. Quantos crimes, guerras, assassinatos, miséria e horrores não teria evitado ao gênero humano quem, arrancando as estacas e preenchendo os buracos, tivesse gritado a seus semelhantes: ‘preserve-os de escutar a esse impostor! Estareis perdidos se esquecerdes que os frutos são de todos e a terra não é de ninguém (*ibidem*).

Rousseau, de fato, parece ter sido o mais antropológico dos contratualistas modernos. Ele previa um ponto de passagem exato que tiraria o

homem da sua condição original, em que nada lhe faltava, não havia preocupação com o futuro e as desigualdades naturais eram irrelevantes. Entre o homem primitivo e a sociedade civil, há uma sucessão de eventos, como o advento do ócio, a expansão das famílias e dos sentimentos de vaidade, o orgulho que colocou fim a inocência humana. Para percorrer esse caminho, o gênero humano passou por sensíveis “revoluções”: “foi essa a época de uma primeira revolução que deu lugar ao estabelecimento e a diferenciação das famílias, e introduziu uma espécie de propriedade; disso talvez nasceram muitas querelas e combates” (idem, p.310).

O estímulo à linguagem somado a um tempo ocioso que o gênero humano não tinha à sua disposição quando vivia livre, conforme bem lhe aprouvesse e distante dos demais, fez com que surgisse o desejo de comodidades desconhecidas aos seus ancestrais; uma primeira submissão à própria vontade, imposta sem que se dessem conta. O trabalho simples de estender o braço para alcançar outra fruta, caso alguém lhe tomasse, foi sendo substituído paulatinamente por sentimentos de inveja e vaidade que não faziam parte da história das gerações anteriores. A metalurgia e o refino das ferramentas em concomitância com o melhoramento dos processos de cultivo da terra, para produzir estoque excedente para atender o futuro e não apenas o presente de suas necessidades, puseram fim definitivo ao espírito humano livre e despreocupado de outrora. Veio aquilo que Rousseau chama de segunda revolução:

A metalurgia e a agricultura foram as duas artes cujo invento produziu esta grande revolução. Para o poeta são o ouro e a prata. Mas para o filósofo são o ferro e o trigo que civilizaram aos homens e que perderam o gênero humano; um e outro eram desconhecidos dos selvagens da América que por isso permaneceram sempre tais; os demais povos parecem, inclusive, ter permanecido bárbaros enquanto praticaram uma destas artes, sem a outra; e talvez uma das melhores razões pelas quais a Europa tenha sido, se não antes ao menos mais contemporaneamente melhor civilizada que as demais partes do mundo, é por

ser mais abundante em ferro e mais fértil em trigo (idem, 317s).

Por se tratar de um filósofo da moral, Rousseau vê na filosofia política contratualista um tipo de normatividade específica. Ele não sugere que o gênero humano possa regredir no tempo e resgatar a liberdade absoluta, perdida por conta de sucessivos avanços que alteraram as relações entre os homens de forma absoluta. Faz parte do projeto do contratualismo, no entanto, observar esse panorama descritivo da natureza humana, para nortear a vontade individual para que ela coincida com a vontade coletiva. A desigualdade que os homens adquirem por conta do advento da sociedade civil, nesse sentido, não é substancial, mas gera desentendimentos nas vontades individuais, produzidos por um simulacro de diferença de poder. No que se refere às suas vontades, os homens são genuinamente iguais, suas necessidades básicas são as mesmas, quando descortinada sua natureza humana simples, avessa ao contato com os outros e livre.

Rousseau rompe a hegemonia de um contratualismo que outorga direitos naturais. Não há, para ele, a necessidade de renúncia da liberdade para o estabelecimento do contrato e novos direitos. Os indivíduos, em condição de livres, são desprovidos de direitos, só adquiridos quando já cidadãos de um Estado. Os homens têm seus direitos submetidos aos direitos supremos de uma comunidade. A igualdade que é relevante, em meio ao contrato social, é aquela estabelecida por convenção e direito legal. Muito mais do que se configurar como parte da realidade, o contrato originário é uma forma de legitimar o vínculo social entre os homens.

4. Liberalismo: a liberdade é um valor político fundamental?

O problema da liberdade é uma questão filosófica genuína, e saber se estamos lidando com um valor político não é tão simples assim. Em primeiro lugar, precisamos deixar de lado muitas percepções do conceito de liberdade também influentes na seara filosófica. Em segundo lugar, existe a necessidade de compreender que a aceitação do conceito de liberdade no ambiente da filosofia política envolve uma percepção ligada tanto ao grau de liberdade que se propõe quanto à forma que esse conceito tem, na sua possibilidade de se efetivar

ou não dentro de diferentes comunidades políticas.

4.1. John Stuart Mill

Um dos primeiros a tratar do problema da liberdade como um valor genuinamente político foi John Stuart Mill (1806-1873). Mesmo que a tradição liberal não tenha iniciado propriamente com ele, sua posição foi a mais influente na filosofia política desde então, especialmente porque sua obra *A Liberdade* é declaradamente uma investigação “pura” da liberdade política. Já nas primeiras palavras do texto, ele “limpa terreno”, mostrando-se desinteressado em um conceito amplo de liberdade, a chamada liberdade metafísica:

O assunto deste ensaio não consiste na assim chamada Liberdade do Arbítrio, oposto de modo tão infeliz à doutrina inadequadamente designada de Necessidade Filosófica, mas na Liberdade Social ou Civil, ou seja, a natureza e os limites do poder que a sociedade pode legitimamente exercer sobre o indivíduo (MILL, 2000, p.5).

O anúncio de Mill deixa de lado boa parte das perspectivas históricas de investigação em torno do problema da liberdade, no seu sentido metafísico, *freedom*. Saber se existe ou não fundamento em sustentar uma visão de mundo determinista, se há ou não algo como destino, não faz parte da agenda da filosofia política a partir da modernidade, preocupada com uma liberdade política, *liberty*. Nesse aspecto, a pressuposição é sempre a favor da liberdade. Quando investigada a partir da liberdade política, essa pressuposição recebe o nome de liberalismo.

A corrente de filosofia política inaugurada por Mill, mesmo delimitada, ainda possui uma gama de respostas muito variadas para a pergunta “a liberdade é um valor político fundamental?” Para Mill, definitivamente, é. A liberdade é uma pressuposição política, no sentido de que o ônus da prova de se uma determinada liberdade política deve ser suprimida ou não está sempre depositado naquele que defende a supressão. A pressuposição *a priori*, antes de qualquer experiência, sempre favorece a liberdade política.

Foi nesse sentido que Mill defendeu, por exemplo, o direito de voto para

as mulheres: aqueles que não concordavam com a ampliação do direito de sufrágio feminino deveriam apresentar argumentos que sustentassem a impossibilidade de as mulheres terem o direito ao voto. O trabalho de Mill no séc. XIX foi a principal referência para a garantia dos direitos que as mulheres vieram a ter na Inglaterra no início do séc. XX.

É errado pensar que a liberdade é apenas um valor político fundamental para um autor como Mill: ele reconhece uma esfera que não é política do conceito. Seu liberalismo, nesse sentido, é político na medida que busca resguardar liberdades inalienáveis que não podem ser pautas de discussão na esfera decisória da política. É dele a famosa expressão “sobre si mesmo, sobre seu corpo e mente, o indivíduo é soberano” (MILL, 2000, p.18). Nesse sentido, existe um espaço da liberdade alheio ao ambiente de discussão da filosofia política. Mill não acredita que seja um problema a ser discutido no Congresso Nacional a sexualidade dos cidadãos ou a predileção das pessoas a serem beneficentes, por exemplo, mas se há direitos políticos envolvidos o assunto torna-se pertinente. Vejamos o exemplo do casamento entre pessoas do mesmo sexo, algo que agrega direitos civis aos envolvidos: Mill exigiria, nesse caso, que o impedimento da união entre os envolvidos tivesse uma justificação publicamente compartilhada, pois, a princípio, a esfera da liberdade política não pode impedir o acesso a esse direito a ninguém em condições de exercê-lo. Aqueles contrários a união civil entre pessoas do mesmo sexo são integralmente responsáveis por justificar os motivos contrários a esse tipo de política pública. O problema enfrentado nessa justificação é o mesmo que foi colocado em pauta quando Mill advogou o direito de voto para as mulheres: o que torna essas pessoas politicamente diferentes das outras para lhes negar acesso ao direito político? Quem for contra precisa responder essa pergunta, pois, fora do certame político, tendo a liberdade como um valor fundamental, não há espaço para interferências na vida dos indivíduos, sem boas justificativas.

4.2. Isaiah Berlin

Isaiah Berlin (1909-1997) também foi um liberal. No entanto, ele defende aquilo que ficou conhecido como “liberdade negativa”, ligada restritamente à ausência de impedimentos. Para ele, o ponto central da liberdade é ausência de coerção pelos outros; principalmente coerção Estatal, embora não

apenas ela. Nesse sentido, o principal papel do Estado é garantir a liberdade, impedindo que os cidadãos coajam uns aos outros, ao menos sem uma boa justificativa para isso. A liberdade é valiosa para Berlin, na medida em que podemos saber quais opções nos são deixadas em abertas, independentemente de seguirmos ou não tal ou tal opção.

Vale notar que a liberdade negativa não é o mero impedimento de realizar alguma atividade: um indivíduo cego, por exemplo, não pode dizer que sua liberdade está sendo obstruída porque ele não pode ler. Berlin se refere a um impedimento deliberado, praticado por outro ser humano em um espaço que estaria aberto para a realização, caso o impedimento não fosse praticado.

Eu, normalmente, sou livre no estágio no qual nenhum homem ou corporação de homens interfere na minha atividade. Liberdade política nesse sentido é apenas a área na qual um homem pode agir sem obstrução de outros. Por outro lado, se eu estou sendo impedido por outros de fazer o que eu poderia fazer, estou no estágio da ausência de liberdade; e se essa área é afetada por outros homens para além de certo mínimo, eu posso ser apontado como sendo coagido ou, talvez, escravizado (BERLIN, 1969, p.122).

Apesar de ser possível argumentar que Rousseau já defendia uma perspectiva de “liberdade positiva”, é com Immanuel Kant que temos uma noção bem acabada do que se trata esse conceito para a filosofia política.

4.3. Immanuel Kant

Immanuel Kant (1724-1804) pensava que uma pessoa que age livremente é autogovernada, pensa por si mesmo sobre as alternativas à sua disposição no mundo. Alguém é livre de acordo com o grau que efetivamente se autodetermina, e o valor político da liberdade é sempre contraposto à escravidão com relação aos próprios desejos. Kant chama essa emancipação de si mesmo de “esclarecimento”, a capacidade de reconhecer, através da própria razão, as

leis universais que devem nortear a vida pública dos cidadãos.

Para os pensadores que seguiram essa linha kantiana, a liberdade é frequentemente apresentada como o poder de agir efetivamente ou perseguir determinados fins por conta própria. A liberdade só é um valor político a ser levado em conta quando estamos diante de um agente moralmente livre que reconhece as regras da sociedade como imperativas, obrigatórias, mas não por força da coerção, e sim por uma compreensão reflexiva, interna a si mesmo. O agente político kantiano reconhece nas leis da sociedade sua própria vontade e, conseqüentemente, a de seus concidadãos. Nesse sentido, Kant escreve: “para este esclarecimento, porém, nada mais se exige senão LIBERDADE. E a mais inofensiva entre tudo aquilo que se possa chamar de liberdade, a saber: a de fazer *uso público* da razão em todas as questões” (1784, p.2). A razão humana deve fazer uso do espaço político de forma autônoma, e aquilo que há de valioso na concepção de liberdade de Kant é tratar todos os seres racionais como capazes de se guiar por uma vontade universal, descoberta por si mesmo, idêntica, portanto, à sua própria vontade.

4.4. John Rawls

John Rawls (1921-2002) foi herdeiro direto da tradição kantiana, e estava disposto a corrigir aparentes problemas do legado deixado pelo contratualismo. Para ele, tanto Mill quanto Kant esbarravam no empecilho de tratar a liberdade a partir de um valor pré-determinado por teorias morais, respectivamente, utilitarismo (tendo em vista as conseqüências das ações) e deontologia (uma ética do dever). Isso se apresenta como um problema sério na obra *Uma Teoria da Justiça*, porque Rawls acredita que precisamos de uma teoria política que não sofra influência daquilo que ele chamou de “teorias morais abrangentes”. Para ele, não é possível convencer as pessoas que elas devem agir de uma determinada maneira moral, para, depois, colher os resultados desse comportamento nas sociedades políticas. Em outras palavras, Rawls queria uma teoria política desvinculada de uma teoria do valor.

Em virtude dessa imposição em torno de precisar escolher por um ou outro modelo de normatividade, ele emprega uma crítica fundamental às teorias políticas que necessitam de uma valoração da moral para realizar seus objetivos. Nota-se isso nas palavras dirigidas contra os utilitaristas:

Deve-se supor uma certa semelhança entre os indivíduos, por exemplo, a sua capacidade de desenvolver atividades e interesses humanos na qualidade de seres capazes de progresso, e além disso o princípio do valor marginal decrescente dos direitos básicos, quando atribuídos aos indivíduos. Na falta desses pressupostos, a promoção dos objetivos humanos sempre pode ser compatível com a opressão ou, pelo menos, com a severa restrição da liberdade de alguém (RAWLS, 2002, p.228s).

Para resolver esse problema, Rawls estava convencido de que era preciso partir de pressupostos novos, desconectados de conceitos de liberdade valorativos, mas igualmente liberais. Ele projetou sua teoria para que servisse como um “constructo teórico” em torno das instituições de justiça da sociedade. O objetivo do liberalismo político não deve ser o de “adicionar” às teorias morais abrangentes outras doutrinas morais sobre o comportamento humano. Religiões podem ser consideradas tão eficientes para normatizar o comportamento das pessoas quanto uma teoria moral kantiana ou uma utilitarista. Não deve ser papel da filosofia política advogar em torno dessas perspectivas. Para Rawls, a filosofia política liberal deve se ocupar de fornecer uma “moldura neutra”, alheia a teorias morais controversas entre si. O liberalismo deve ficar restrito a um conjunto básico de princípios políticos que possam ser tomados como consenso entre todos os cidadãos. Só assim é possível o estabelecimento de instituições políticas que respeitem uma gama muito diversificada de agentes morais.

Rawls levou a sério a possibilidade de construir uma sociedade plural com uma teoria política liberal austera. Ele desconectou o liberalismo político das teorias morais e o colocou ao lado de uma teoria do direito. A construção de seus princípios políticos dependia exclusivamente de um acordo coletivo – portanto, contratualista – para estabelecer as instituições a partir de uma deliberação coletiva por trás do chamado “véu da ignorância”. Trata-se de um experimento teórico que busca fazer com que os participantes do jogo político, responsáveis pelo estabelecimento da constituição, se coloquem na posição de legisladores capazes de deliberar imparcialmente, sem partir de nenhum pressuposto sobre

suas condições materiais reais na sociedade em que vivem.

Temos, em Rawls, uma forma de liberalismo igualitário que leva em consideração dois princípios de justiça: o primeiro prescreve liberdades fundamentais que devem ser iguais para todos; o segundo trata de estabelecer um critério para distribuição dos recursos disponíveis que prioriza o atendimento aos menos favorecidos. Para Rawls, não se trata de um nivelamento igualitário dos recursos da sociedade, mas equitativo, com igualdade de oportunidades; e sua preocupação é justamente com a estabilidade política de um ambiente tão plural. Ele quer saber como pode perdurar no tempo uma sociedade estável e justa, de cidadãos livres e iguais, mesmo com uma profunda divisão em doutrinas religiosas, filosóficas e morais. Para Rawls, o liberalismo é uma concepção política que, mesmo sendo independente de qualquer doutrina abrangente e racional (metafísica ou epistemológica), busca um “consenso sobreposto” por parte dos envolvidos. Trata-se de uma base pública de justificação com características contratualistas, universalmente aceitável pelos cidadãos e possível de ser empregada em sociedades liberais que toleram concepções de bem razoáveis em sua esfera pública de decisão, independente do credo filosófico, moral ou religioso.

O liberalismo igualitário de Rawls tornou-se uma resposta aparentemente funcional se levarmos em conta as exigências de uma sociedade bem organizada, com base estabelecida na justiça e com um amplo espectro de pluralismo.

Rawls não negava uma sociedade com diferenças entre ricos e pobres, mas apostava ser possível – ao menos em teoria – construir instituições de justiça suficientemente plurais para dar a todas as concepções de bem que fossem razoáveis espaço político para desenvolver seus projetos de vida. A distribuição de condições equitativas para os membros dessa sociedade é parte do processo de construção, a chamada igualdade de oportunidades. Não se trata, portanto, de tornar a vida das pessoas materialmente iguais, mas de respeitar as diferentes concepções de bem que possam existir na sociedade liberal, para que todos possam levar adiante seus projetos de vida com condições equitativas de oportunidade. A “igualdade de oportunidades” sugere que quaisquer outras desigualdades que ocorrerem devido às escolhas voluntárias dos agentes serão justas, desde que dentro de um cenário de igualdade de oportunidades.

Questões para revisão

1. Maquiavel influenciou o renascimento com seu republicanismo. Para ele, a política é uma ciência de observação dos defeitos e correções do Estado. Explique esse novo tipo de realismo político. Qual o papel da *virtù* e da *fortuna* em meio à análise da história dos Estados?
2. O bem-estar dos cidadãos independe da tripartição tradicional das diferentes formas de governo. Montesquieu está preocupado em ampliar a liberdade da sociedade civil, pulverizando o governo em diferentes poderes internos ao seu funcionamento. Explique o que precisa ser limitado para garantir um governo melhor e mais eficiente, para esse autor.
3. Agostinho queria ver a Cidade de Deus operando entre os homens na terra. Qual sua inspiração e motivação para pensar esse tipo de poder político? Explique.
4. A referência de Tomás de Aquino foi Aristóteles. Ele tentou conjugar o poder divino, revelado, com uma explicação racional. Qual o resultado disso para o poder sacerdotal?
5. João de Quidort sugeriu uma separação entre os poderes do Estado e a Igreja. Como e por quê?
6. O contratualismo foi retomado com força teórica por Hobbes. Foi uma hipótese de trabalho ou ele acreditava em um contrato originário, entre os homens para formação do Estado? Explique as principais partes do argumento.
7. Locke deu um peso diferenciado e materialista para a ideia de propriedade privada no contratualismo. Explique.
8. Para Rousseau, a desigualdade da sociedade surge de uma compreensão inadequada da noção de propriedade. Qual deveria ser o foco do contrato original?
9. Mill apresentou uma expectativa política de liberdade, enquanto valor das comunidades políticas contemporâneas. Qual foi o critério dele?
10. Explique a diferença entre liberalismo negativo (Berlin) e esclarecimento (Kant).
11. O liberalismo político de Rawls é desconectado de uma concepção valorativa tradicional. Fale da importância do construtivismo e do véu da ignorância para instituições que promovem a justiça.

Questão para discussão

1. Platão e Aristóteles delimitaram o problema político da melhor forma de governo de maneiras diferentes, ambas influenciaram toda a história da filosofia. Discorra sobre a diferença básica de ambas as teorias, o governo ideal e o bom governo ou estável.

Referências

- ABBAGNANO, Nicola. *Dicionário de Filosofia*. Trad. Alfredo Bossi. 5ª ed. Martins Fontes: São Paulo, 2007.
- ARISTÓTELES. *Política*. Trad. Antônio Campelo Amaral e Carlos de Carvalho Gomes. Vega: Porto, 1998.
- BAVARESCO, Agemir; SANTOS, João V. F. dos; KONZEN, Paulo R. O Princípio da Utilidade e da Liberdade de Expressão e de Informação, em Sobre a Liberdade, de J. S. Mill. *Agora Filosófica*. Recife: pp.7-47, Unicap, n.1, 2012.
- BENTHAM, Jeremy. *Uma Introdução aos Princípios da Moral e da Legislação*. Trad. Luiz João Baraúna. Coleção Pensadores, v.34. São Paulo: Abril Cultural, 1974.
- BERLIN, Isaiah. *Four Essays on Liberty*, Oxford University Press: Oxford, 1969.
- _____. Introdução. In MILL, John Stuart. *A Liberdade; Utilitarismo*. Trad. Eurice Ostrensky. São Paulo: Martins Fontes, 2000, pp. VII-LIV.
- BOBBIO, Norberto. *A Teoria das Formas de Governo na História do Pensamento Político*. Trad. Luiz Sérgio Henriques. Edipro: São Paulo, 2017.
- HOBBS, Thomas. *Leviatã: ou matéria, forma e poder de um Estado eclesiástico e civil*. Trad. Daniel Moreira Miranda. Edipro: São Paulo, 2015.
- HUME, David. *História da Inglaterra: da invasão de Júlio César à Revolução de 1688*. Trad. Pedro Paulo Pimenta. 2ª ed. Unesp: São Paulo, 2017.
- KANT, Immanuel. *Resposta à Pergunta: Que é esclarecimento?* (1784), acessado em jul. 2020 <<https://everleit.files.wordpress.com/2013/05/kant.pdf>>.
- LADNER, Gerhart B.. Aspects of Mediaeval Thought on Church and State In *The Review of Politics*, Vol. 9, No. 4, pp.403-442. Cambridge University Press: Cambridge, Oct., 1947.
- MAQUIAVEL, Nicolau. *O Príncipe*. Trad. Maria Júlia Goldwasser. 4ª ed. Martins

- Fontes: São Paulo, 2010.
- _____. *Discorsi*: comentários sobre a primeira década de Tito Lívio. Trad. Sérgio Bath. 3ª ed. Editora da Universidade de Brasília: Brasília, 1994.
- MILL, John Stuart. *A Liberdade, Utilitarismo*. Trad. Eunice Ostrensky. Martins Fontes: São Paulo, 2000.
- PIERPAULI, José Ricardo. Las Relaciones entre el Poder Eclesiástico y el Poder Político los Casos de Alberto Magno, Tomás de Aquino, Juan Quidort, Marsilio de Padua y Alvaro Pelagio. In *Dissertatio* n. 29, pp.115-133. UFPel: Pelotas, inverno de 2009.
- PLATÃO. *A República*. Trad. Maria Helena da Rocha Pereira. Calouste Gulbenkian: Lisboa, 2001.
- RAWLS, John. *Uma Teoria da Justiça*. Trad: Almiro Pisetta; Lenita M. R. Esteves. 2. ed. São Paulo: Martins Fontes, 2002.
- ROUSSEAU, Jean-Jacques. Discurso sobre el Origen y los Fundamentos sobre la Desigualdad entre los Hombres. Trad. Francisco Márquez Cabrera. In *Obras Selectas*. Madri: Edimat, 2000.
- SILVA, Lucas Duarte. Os Argumentos de Francisco de Vitoria Contrários à Teoria da Escravidão. In *Natural Razão e Fé*. UCPel: Pelotas, 15(2): pp.57-76, Jul.-Dez., 2013
- SIMÕES, Mauro C. Utilidade e Liberdade em John Stuart Mill. *Enfoques*. Buenos Aires: Universidad Adventista del Plata, v.27, n.1, out., 2005, pp.77-83.
- TOMÁS DE AQUINO. *De Regno*. Trad. Francisco Benjamin de Souza Neto. Editora Vozes: Petrópolis, 1997.

Filosofia da Economia

Ramiro de Ávila Peres

1. Introdução

O escritor A. J. Jacobs (2018) decidiu agradecer todas as envolvidas na preparação de seu cafezinho diário; isso o levou a uma jornada épica para contatar milhares de pessoas, desde a barista que lhe vendia o expresso matinal, passando por designers e inventores, até os agricultores que plantaram e colheram o grão em outro continente. Não existe um único indivíduo ou grupo responsável pelo seu café: é preciso um *mundo* para produzi-lo; é resultado de uma cadeia produtiva, uma *rede descentralizada* que envolve diversos serviços espalhados internacionalmente – e todas essas pessoas, por sua vez, consomem outros produtos, com outras cadeias. Como este livro.

Contraste esse exemplo com a anedota do diplomata soviético que, visitando Londres, fica maravilhado com as padarias – e então pergunta pelo encarregado da distribuição de pães na cidade (Harari, 2016, p. 372). Para seu espanto, os economistas ingleses respondem que ninguém tem esse cargo: o fornecimento de pães é resultado de um agregado de decisões individuais, em que cada fornecedor de matéria-prima (como farinha, sal, fermento) e serviços (transporte, padaria, limpeza) regula sua *oferta* a partir do que espera que os consumidores querem consumir – e o quanto estes estão dispostos a pagar. É o caso clássico do funcionamento de uma *economia de mercado* – em contraste com uma *economia de comando*, onde um grupo de pessoas, conscientemente, toma decisões sobre produção e distribuição.

Esses exemplos ressaltam as “vantagens” dos mercados; pode ser tentador pensar, a partir deles, que todos os bens e serviços deveriam ser fornecidos dessa forma, porque isso maximiza a liberdade individual ou tende a

levar ao bem-estar coletivo, e que o melhor a fazer é evitar interferências nesse sistema – ideias associadas ao liberalismo econômico, uma teoria política e econômica que tem por patrono Adam Smith (o primeiro economista clássico). No entanto, mesmo economistas liberais reconhecem que há *falhas de mercado* (como bens públicos e monopólios¹²⁹). Voltemos a A. J. Jacobs, o qual frisa seu agradecimento ao serviço de fornecimento de água, um “milagre moderno”: ao longo da história, a imensa maioria das pessoas não teve acesso a água encanada – mesmo hoje, 2 bilhões de pessoas ainda carecem de condições sanitárias adequadas (água e esgoto), o que causa doenças que matam 829 mil por ano (WHO, 2015). É um problema que não pode ser resolvido só por mercados, nem por associações filantrópicas; na maior parte do mundo, é um serviço público, geralmente fornecido ou regulado por estados – e sua ausência decorre de falhas de governo.

Mercados e governos são instituições, e podemos desenhá-los de formas melhores ou piores – a fim de evitar situações em que, agindo em prol do próprio interesse, chegamos a um resultado pior para todos. Por exemplo, nossas sociedades gastam mais com refrigerantes do que com testes para Covid-19, e mais com esportes de inverno do que em mitigação de riscos

¹²⁹ Bens públicos são bens *não rivais* e *não excludentes* (Cowen & Tabarrok, 2013, p. 345) – p. ex., você pode usufruir de ar puro sem reduzir sua disponibilidade para os demais (rivalidade), e você não pode impedir que outros também o usem (não legalmente). Isso impede a alocação privada desses bens e o estabelecimento de um mercado; isso também dificulta distribuir o ônus do financiamento de tais bens: imagine que alguém invente uma tecnologia que custa \$10 para limpar o ar de nossas cidades; se eu souber que você está disposto a pagar os \$10, posso dizer que não vou pagar nada, pegar “carona” e aproveitar o ar puro de graça. Monopólios, por sua vez, são situações em que um único agente é o fornecedor de um serviço – seja por disposição legal (p. ex., concessionárias de serviço público ou propriedade intelectual), ou por condições do mercado (um monopólio “natural”, quando uma firma pode suprir todo o mercado a um preço menor do que duas, graças a economia de escala e rede). Como o monopolista concentra o poder no respectivo mercado, ele pode determinar os preços de forma unilateral, e assim “extrair renda”, ganhando muito mais do que se estivesse num mercado competitivo; mas isso pode levar a uma quantidade de vendas (e de bem-estar total) menor – o que Cowen & Tabarrok (2013, p. 240) chamam de “peso morto” do monopólio.

catastróficos globais; isso é um resultado de nossas decisões individuais sobre gastos, mas não significa que, no geral, as pessoas valorizem mais o suprimento de bebidas do que prevenir e mitigar pandemias. Críticos dos mercados (principalmente em finanças) também costumam apontar sua instabilidade; p. ex., mercados financeiros podem ter variações abruptas, e cadeias produtivas eficientes podem colapsar rapidamente e demorar para se recuperar – especialmente no caso de desastres.

Por outro lado, defensores do liberalismo econômico destacam que, conquanto mercados não solucionem problemas morais por nós, eles resolvem questões práticas de alocação de recursos; e o correspondente desenvolvimento econômico tem sido fundamental para a melhoria das condições de vida desde a Revolução Industrial – e inclusive teria tido como “efeito colateral” o progresso moral de nossas sociedades, ao tornar cada vez mais obsoletas formas tradicionais de divisão de trabalho, exploração e conflito, sugerindo a possibilidade de um futuro ainda melhor (Cowen, 2018).

2. A importância do desenvolvimento econômico e a complexidade da economia

Imagine como seria viver como um caçador-coletor nômade no final do período Paleolítico (mais de 12 mil anos atrás): passar o dia fazendo *trekking* por florestas e prados com seus colegas de tribo, aproveitando a caça abundante, com as flechas que você mesmo fez ou que ganhou de presente, sem grandes preocupações. À noite, festas e histórias ao redor da fogueira, sob o céu estrelado; todos se conhecem, e não há desigualdade: o líder não é uma classe separada, mas apenas o primeiro entre os pares (veja uma descrição interessante no primeiro capítulo de Harari, 2015).

Há um *viés do sobrevivente* nessa figura idílica: concentramo-nos apenas na amostra que passou por um processo de seleção, ignorando os que não superaram tal processo. Embora haja enfatizado a ausência de desigualdade, não mencionamos que, p. ex., seu papel na vida seria definido pelo seu gênero. Também não contamos o irmão que você perdeu graças à alta mortalidade infantil; nem o tio que morreu num atrito com uma tribo inimiga, que vocês massacraram – e que havia se deslocado para seu território graças a uma estiagem causada pelo clima, ou a alguma doença misteriosa. Nem o avô que

morreu aos 40 anos de uma doença desconhecida, nem seu primo que teve septicemia ao cortar um dedo, ou seu amigo que foi ostracizado depois de comer algo sem fazer o ritual adequado.... A vida no paleolítico era muito boa para os afortunados que *sobreviviam*; para os demais, era brutal (Pinker, 2011, 74).

Para ser justo com esse período, a maior parte dos 12 mil anos seguintes apresentou problemas similares para a maioria das pessoas. Mas algo mudou “abruptamente” no século XIX: a Revolução Industrial levou a um aumento da produção, melhorando o nível de vida dos países industrializados – que passaram a investir, p. ex., em saneamento, pesquisa e educação (Rosling, Rosling & Rönnlund, 2019). Então essa mudança se transferiu para o resto do mundo: a expectativa de vida mundial até o começo do século XX era de cerca de 31 anos – hoje, **é de 72 anos**; e a população aumentou de 1 para 7 bilhões de pessoas.

A explicação dessa mudança é que passamos de sistemas econômicos primitivos baseados na subsistência (você fazia suas próprias flechas e caçava) para economias de escala, com complexas divisões de trabalho baseadas na especialização, no comércio e na inovação. Lembre-se do exemplo do café de A. J. Jacobs: há pesquisadores e designers tentando tornar a “experiência” de tomar café mais agradável e mais barata, agricultores plantando variedades selecionadas para um melhor sabor ou adaptação ao clima, há toda a cadeia de produção e entrega do produto... e, claro, há as pessoas que trabalham para melhorar a vida desses trabalhadores (como prover água tratada), ou para mitigar efeitos negativos que essas atividades possam ter sobre outros (p. ex., reduzir a sua pegada de carbono).

Isso tem consequências morais e políticas: a interrupção do desenvolvimento econômico implicaria maior sofrimento (devido à diminuição das perspectivas de vida) da população atual e futura, e possível aumento da desigualdade e da instabilidade política. Mas então seria fundamental entender melhor como o desenvolvimento ocorre; p. ex., por que a revolução industrial ocorreu precisamente na Inglaterra, no fim do século XVIII – e não na Europa continental, onde havia uma elite científica e cultural igualmente avançada? Por que o desenvolvimento econômico se espalhou pelo mundo no século XX, da maneira como ocorreu? Não temos respostas precisas a essas perguntas; os economistas concordam sobre um conjunto de fatores, mas não sobre as relações causais precisas entre eles. É um pouco como se um paciente tomasse

um coquetel de drogas, sem haver consenso sobre a *interação* entre elas – ou, mais modernamente, podemos relacionar a expressão de genes a uma determinada doença, sem determinar precisamente como ela é causada.

Entre as causas da aceleração do desenvolvimento econômico, Acemoglu e Robinson (2013) destacam a estabilidade política, a democracia, o estado de direito e a ausência de instituições *extrativistas* (i.e., que permitem que uma elite ou oligarquia extraia renda dos demais, como regimes coloniais); outros enfatizam a liberalização econômica (que leva à destruição criativa: a substituição de empresas improdutivas pela concorrência), a educação, que aumenta a produtividade do trabalho e incentiva a pesquisa (Lisboa & Pessoa, 2020, p. 412), ou uma política industrial adequada. Mas há fatores altamente contingentes, como geografia: p. ex., note como quase todos os países desenvolvidos estão em climas temperados; Dell, Jones e Olken (2012) encontraram uma correlação negativa entre crescimento econômico e calor a partir de dados do último meio século – sugerindo que, com o aquecimento global, a produtividade econômica em geral poderia declinar. Isso implica não apenas que há fatores difíceis de controlar, mas também que mudanças climáticas teriam consequências difíceis de prever sobre a economia.

Quadro 1: Um modelo de economia moderna

Uma economia moderna tem um “setor de produção”, populado por firmas. Estas são compostas por pessoas com diferentes interesses (*stakeholders*), como: a) trabalhadores com divisão de tarefas especializadas em linhas de montagem (como operários, montadores, vendedores, contadores, etc.); b) investidores (que financiam as atividades da firma, no presente, esperando obter lucros futuros – como credores); c) clientes e consumidores (que pagam a firma por seus produtos e serviços); d) gestores e executivos (uma classe especial de trabalhadores – as pessoas que tomam as decisões em nome da firma); e) sócios e acionistas (uma classe especial de investidores – são os “donos” da firma, que aceitam uma parcela maior de risco e de lucro, mas em troca têm o poder de controlar a atuação dos gestores).

Um outro setor importante é o governo / Estado. Política e juridicamente, costumamos dividi-lo em *poderes*: o Judiciário (que tem a palavra final sobre a aplicação do direito), o Legislativo (que representa a sociedade, cria leis e fiscaliza outros poderes) e o Executivo; neste que se concentram a maioria das decisões e atividades do governo: a taxação (financiamento das atividades do governo mediante tributos), finanças (decisões sobre gastos e investimentos com impacto macroeconômico, como o endividamento público, a taxa de juros, etc.), redistribuição e seguridade social (pagamento de benefícios sociais) e prestação de serviços públicos (como segurança, saúde pública, transporte e infra-estrutura).

Não são indivíduos, são *posições* – que podem ser ocupadas inclusive por um mesmo indivíduo. Por exemplo, você pode ter empréstimos (consumidor) e depósitos (investidor) num banco, comprar suas ações (sócio), trabalhar para ele (trabalhador), subir na carreira e virar um executivo. Além disso, algumas firmas podem ter divisões de trabalho bastante diferentes, como cooperativas e *start-ups*. Mas, no geral, mesmo essas organizações adotam modelos mais hierarquizados à medida que se desenvolvem, levando a um setor produtivo dominado por instituições hierarquizadas. Assim como Estados hierarquizados, com centralização de decisões e divisão de tarefas, teriam surgido a partir de conflitos violentos entre populações, também o modelo de firmas hierarquizadas é selecionado pela competição numa economia de mercado (Coase, 1960).

Há muitos exemplos de intersecções proveitosas entre o estudo da

Filosofia e da Economia¹³⁰. Por questões de espaço (pensa nisso como uma lição sobre custos de oportunidade), e para não repetir temas que você viu em outros capítulos (como em Ética e Filosofia Política), focaremos os temas a seguir, tendo como pano de fundo assuntos associados à área de Finanças – um setor cujo funcionamento é pouco conhecido por não-especialistas, a despeito de sua importância para o processamento de informações e alocação de recursos em sociedades capitalistas (Soros, 2011).

- 1) *O problema da incerteza* em economia: a epistemologia da economia oferece desafios peculiares. Não apenas se buscam desvendar padrões e leis gerais (o que já é suficientemente difícil) sobre a atividade econômica, mas também antecipar o que outros agentes farão; mas estes juízos factuais são altamente inseguros. Um dos papéis importantes dos economistas é desenvolver teorias e modelos matemáticos complexos para explicar e *prever* eventos, além de desenhar instituições para o processamento de informações economicamente relevantes (como mercados).
- 2) *Moeda, legitimidade e justiça*: vamos discutir como as notas coloridas que você guarda na carteira tem uma relação com a distribuição de poder político e econômico na sociedade. Devemos reduzir essa concentração de poder ao máximo (em nome da liberdade individual – como hoje propõem defensores das criptomoedas), diluí-lo entre os cidadãos (em nome da democracia) ... ou devemos outorgá-lo a tecnocratas apartados do processo político – que é como as coisas de fato funcionam?

¹³⁰A começar pela própria definição de Economia – uma ciência que inclui desde a análise teórica da tomada de decisão (como enfatizado pela literatura em teoria da firma, teoria dos jogos, teoria da escolha social...) até o estudo da organização, produção e do desenvolvimento em sociedade. Outras omissões importantes: discussões sobre regimes de direitos de propriedade (incluindo propriedade intelectual, de empresas ou de bens coletivos), agência moral corporativa (devemos tratar empresas como moralmente responsáveis *per se*, ou apenas como representantes de seus donos?), distribuição e justiça econômica, teoria dos jogos, teoria da escolha social etc.

- 3) *Ética nos negócios*: filósofos frequentemente denunciam a fria amoralidade das decisões econômicas – enquanto economistas apontam que ela decorre do mesmo mecanismo que incentiva a inovação e preços baixos, i.e., o caráter competitivo do mercado. Essa discussão assumiu proporções importantes em tempos recentes, quando diferentes setores produtivos e movimentos sociais passaram a agir em relação a problemas globais, como as mudanças climáticas. Nesse contexto, que tipo de compromissos éticos faz sentido exigir de empresas, além de suas obrigações legais básicas?

3. O problema da incerteza

A Economia é uma ciência que pensa em termos de modelos, junto com a arte de escolher modelos... (Keynes, 1938; apud. Perelman, 2007, p. 169, trad. livre).

Na década de 30, os países industrializados passaram por uma crise econômica sem precedentes, começando pelo *crash* da bolsa de Nova Iorque em 1929 e a Grande Depressão subsequente – um evento que desencadearia reações sociais que culminariam na 2ª Guerra Mundial. No Reino Unido de 1932, mesmo o governo trabalhista falhava em seguir à risca os conselhos do influente J. M. Keynes de aumentar o gasto público e desvalorizar a moeda, a fim de estimular a economia e mitigar os efeitos da crise sobre a classe trabalhadora. Seus amigos, influenciados pela experiência marxista que então se desenrolava na União Soviética, observaram que, se seus conselhos eram tão bons, o governo não teria razão para ignorá-los, exceto para favorecer os interesses da elite financeira. Mas Keynes tinha uma explicação alternativa:

A facção da guerra de classes acredita que se sabe o que deve ser feito; que estamos divididos entre os pobres e bons que gostariam de fazer isso, e ricos e maus que, por razões de interesse próprio, desejam evitá-lo; que os ímpios têm poder; e que é necessária uma revolução para os depor de seus assentos. Vejo o assunto de outra forma. Acho extremamente difícil saber o que

deve ser feito, e extremamente difícil para quem sabe (ou pensa que sabe) convencer os outros de que está certo - conquanto teorias, que são difíceis e obscuras quando novas e não digeridas, se tornem mais fáceis de aceitar com a simples passagem do tempo (Keynes, 1934; apud. Carter, 2020, p. 219, trad. Livre).

Conflitos de interesse e de poder são importantes na economia; no entanto, note que, se a interação econômica não é de cooperação pura, também não é um “jogo de soma zero”, nem de conflito total. Essa é uma importante diferença entre um cassino e o mercado: neste, os investidores não estão apenas fazendo apostas onde “o vencedor leva tudo”; como veremos, eles também provêm informações valiosas sobre bens e ativos, além de liquidez (i.e., recursos no curto prazo) para empresas. O principal problema da economia ainda é o da incerteza das previsões, tanto macro quanto microeconômicas: assim como a crise de 1929, a crise de 2008 (a “Grande Recessão”) foi uma surpresa, como o foi a crise de 2020 (o “Grande Lockdown”). É ‘normal’ que previsões econômicas falhem: desde 1988, o FMI tem feito previsões sobre a possibilidade ocorrer uma recessão (i.e., uma queda no Produto Interno Bruto) num dos 194 países no ano seguinte; das 469 recessões observadas, apenas 17 foram corretamente previstas – além de 47 previsões de recessão que não se confirmaram (Shiller, 2020, p. xiv).

A economia é uma ciência aplicada, que visa a informar a tomada de decisões; mas como escolher racionalmente suas ações, se você não sabe quais serão as respectivas consequências?

4. O “concurso de beleza” de Keynes

O exemplo clássico da peculiaridade do raciocínio microeconômico em mercados é o do *concurso de beleza* de Keynes: imagine que um jornal pede para seu público selecionar os seis rostos mais bonitos entre dezenas de fotografias, prometendo um prêmio para a pessoa que selecionar os seis rostos mais votados. Ao invés de você escolher os rostos que você considera mais bonitos, você usaria o seguinte raciocínio *recursivo*: tentaria adivinhar que rostos os *outros participantes* escolheriam – sabendo que eles também não estão

tentando escolher os rostos mais bonitos, e sim quais os rostos escolhidos por você e pelos demais (Shiller, 2020, p. 63).

Em mercados regidos por um sistema de preços, enfrentamos um problema análogo. Um investidor no mercado de capitais considera sua própria opinião sobre o valor das ações de uma empresa como *uminput* na sua avaliação, sabendo que o preço dessas ações é, na realidade, definido pelo agregado de decisões sobre compra e venda de todos os participantes desse mercado. Cowen & Tabarrok (2013, p. 118) chamam preços de “sinais enrolados em incentivos”: quando você sinaliza que quer comprar ou vender algo por \$10, está, ao mesmo tempo, expressando sua opinião sobre a situação da oferta, da demanda e do custo de oportunidade (i.e., o quanto você poderia ganhar tomando uma decisão distinta) desse bem – incentivando quem acha que ele vale menos a vendê-lo, ou quem acha que vale mais a comprá-lo.

Pensemos num exemplo simples: imagine que Alice acredita que a ação de uma companhia valerá R\$60 na semana que vem, enquanto Bruno acha que esse valor será R\$20; nesse cenário, você pode garantir um lucro sem risco mediante *arbitragem*, prometendo comprar a ação de Bruno, na semana que vem (por mais do que R\$20, mas menos que \$60), e vendê-la a Alice (por mais do que comprou de Bruno)¹³¹. Assim, de acordo com a hipótese dos mercados eficientes (HME), as oportunidades de arbitragem num mercado são rapidamente esgotadas, e os preços acabam por refletir *todas* as informações disponíveis sobre a oferta e a demanda dos produtos e ativos em questão – um resultado epistemicamente superior ao que qualquer agente individual (ou autoridade central) poderia alcançar por si só. Isso implica que o dilema do investidor (“quanto vale esse bem?”) deveria desaparecer: o preço de um ativo já reflete o agregado de juízos sobre seu valor, e a variação subsequente desses preços deveria ser considerada aleatória – donde não seria possível “vencer” o

¹³¹ O mesmo conceito pode ser estendido a apostas: se Alice acredita que um fato tem probabilidade de 60% de ocorrer (e assim estaria disposta a apostar \$60 contra \$40 que o fato ocorrerá), enquanto Bruno pensa em 20% (e apostaria \$10 contra \$40 que o fato não ocorrerá), então você pode oferecer a ambos apostas que eles considerariam justas sem que haja risco algum para você. Se o pagamento ocorrer, você receberá \$40 de Bruno, com o que pagará Alice; se ele não ocorrer, você receberá \$60 de Alice, com o que pode pagar \$10 a Bruno e lucrar \$50. Pessoas que trabalham com derivativos e *hedge* estão à procura de negócios desse tipo.

mercado no longo prazo (i.e., obter lucros significativamente maiores que a média), nem seria possível prever bolhas (pois fazê-lo implicaria “saber mais” do que o mercado) de forma confiável.

Esse é um dos motivos por que economistas frequentemente criticam a intervenção estatal, mediante tributos ou regulação: ela interfere com a maneira como o mercado processa informações e desloca o *equilíbrio*¹³² – e, conseqüentemente, a maneira como mercados transmitem informações relevantes (i.e., o “mercado sabe mais” que a autoridade central). Mas economistas são os primeiros a reconhecer que as condições para aplicação de teoremas de equilíbrio geral raramente estão presentes no mundo real; isso é destacado até mesmo por teóricos do equilíbrio geral, como Kenneth Arrow e Ronald Coase (1991, p. 436). Em especial, restrições *informacionais* (como a seleção adversa, o risco moral e os limites cognitivos dos indivíduos reais) equivalem a custos de transação que não podem ser completamente afastados (Myerson, 2008, p. 131).

Mas note como HME tem algo de paradoxal: assumimos que o mercado é informacionalmente eficiente justamente porque os demais agentes se dão o trabalho de pesquisar e obter informações, que depois serão agregadas pelo mercado, mediante suas respectivas compras e vendas – i.e., eles não tratam a variação dos preços como aleatória; se todos os investidores agissem de acordo com HME, então seria lucrativo para um único investidor pesquisar mais e se antecipar ao mercado (Sorensen, 1988, p. 110). De fato, há investidores reconhecidos por consistentemente obterem retornos superiores ao mercado financeiro e há economistas reconhecidos por preverem bolhas (Weatherall, 2015).

Por fim, note como nosso uso do termo “incerteza” é vago e ambíguo – significando apenas o oposto de informação. Filósofos e economistas frequentemente distinguem a noção de *risco* (eventos negativos com

¹³² Um exemplo quase unânime entre economistas é a defesa do livre mercado e a crítica a tarifas protecionistas: quando um governo levanta barreiras alfandegárias para proteger a indústria nacional, ele incentiva outros países a fazerem o mesmo (para compensar suas respectivas perdas em exportações); mas isso leva a um aumento dos preços em todos os lugares. Portanto, segundo esse argumento, barreiras comerciais só deveriam ser adotadas em situações excepcionais (em caso de crise, ou para fomentar uma indústria nascente), como prega a Organização Mundial do Comércio.

probabilidade precisamente mensurável – i.e., o “risco conhecido”) e de *incerteza* em sentido estrito (i.e., riscos desconhecidos, situações onde não se tem a distribuição de probabilidades do evento – ou onde essa distribuição é estatisticamente anômala). O seu *risco* de perder uma aposta de cara ou coroa com uma moeda imparcial é 50%; mas se você não sabe que a moeda é imparcial, você tem *incerteza*. O seu risco de morte pode ser calculado de acordo com a tábua de mortalidade compilada por atuários; mas qual o risco de você se jogue pela janela? Espero que seja zero – mas esse é um evento incerto cuja probabilidade é afetada por sua disposição e comportamento a todo momento. A HME seria inconsistente com o fato de que o retorno sobre o investimento é dominado por situações raras, que se situariam no limite da distribuição de probabilidades, como uma decisão surpreendente de um governo, uma epidemia, guerras, ou uma nova tecnologia – o que Bremmer e Keat (2009) chamam de riscos de “caudas longas” (*fat-tails*) e Nassim Taleb (2015) chama de *cisnes negros*.

5. Incerteza e a influência das ideias

É reconhecido que a teoria econômica é influenciada pelo contexto histórico das ideias (Arida, 1983); mas isso pode ser extrapolado para outras atividades científicas – vide o capítulo sobre Filosofia da Ciência. Mas, no caso de ciências sociais, a influência é de mão-dupla: ideias e estados psicológicos em sociedade, e as teorias e instituições que construímos para responder-lhes, influenciam resultados econômicos; trata-se de uma diferença importante em relação a outras ciências empíricas: uma partícula pode ser afetada por um “efeito de observador” (i.e., ao medir suas propriedades, o observador precisa interagir com ela), mas ela não “observa de volta”. Soros (2011) chama esse efeito de mão-dupla de *reflexividade*.

Keynes já apontara a correlação entre os ciclos de negócios e o “espírito animal” dos investidores – o que mais recentemente veio a ser chamado de *exuberância irracional* por Alan Greenspan (2013). Mesmo a teoria das expectativas racionais, segundo a qual os investidores antecipam as condições macroeconômicas futuras (p. ex., que o déficit público atual será pago por impostos futuros), apela para *juízos* dos investidores em relação ao porvir. O grande problema é que, se já é difícil compilar, acompanhar e antecipar

estatísticas macroeconômicas, imagine as ideias e modismos que afetam a sociedade.

É justamente o que Robert Shiller (2020) propõe fazer; em seu *Narrative Economics*, ele busca analisar empiricamente a forma como ideias se espalham, usando modelos epidemiológicos e ferramentas de análise textual (p. ex., medindo a frequência de expressões como “Great Recession” no *Google Ngram*). Isso sugere que mesmo alguns padrões fundamentais do estudo da economia podem mudar graças a avanços na tecnologia de transmissão de informações, ou a mudanças nas ideias dominantes (investidores podem responder mais rapidamente a flutuações, ou se tornar mais céticos a respeito da sustentabilidade do endividamento público, por exemplo). Isso também ajuda a explicar como algumas ideias se tornam “profecias auto-confirmatórias” quando são altamente contagiosas.

Por exemplo: em 23 de abril de 2013, o Twitter da Associated Press anunciou uma explosão na Casa Branca – por apenas 5 minutos. Muitos investidores liquidaram suas posições, com o auxílio de sistemas de *trading* de alta-frequência, e o índice Dow Jones despencou, levando a perdas da ordem de US\$ 1 trilhão em valor de mercado (Kirilenko & Lo, 2013). Aparentemente, ninguém realmente levou o anúncio a sério (não há notícia de medo de uma guerra, por exemplo); mas isso não é necessário, pois um investidor minimizará sua exposição a um ativo sempre que atribuir uma probabilidade suficientemente grande à hipótese (H): “um número suficientemente grande de outros agentes vai minimizar suas exposições, e esse ativo vai perder valor” – que se torna assim uma profecia auto-confirmatória.

Há quem argumente que, justamente por causa da grande incerteza que cerca decisões com impacto econômico, autoridades deveriam se abster de agir, pois é melhor errar por omissão que agir por engano - como aduz Huemer (2012) em seu *elogio da passividade*. Isso equivale a uma *objeção cética* contra a legitimidade da intervenção por autoridades técnicas: à direita do espectro político, esse argumento pode ser relacionado a economistas libertários. À esquerda, há um argumento afim em objeções da teoria crítica à tecnocracia: se não há conhecimento confiável sobre o futuro (se a “racionalidade instrumental” falha aqui), então *experts* não estão numa posição privilegiada para tomar decisões sociais – que deveriam, pois, ser tomadas por todos, ou por representantes democraticamente eleitos.

O problema é que, *levada às últimas consequências*, a objeção cética prova demais; se a incerteza nos impedisse de formar estimativas, isso não levaria apenas à impossibilidade do planejamento e da ação social, mas à impossibilidade da ação racional *simpliciter*, pois é impossível agir sem ter expectativas sobre o futuro. A objeção cética lembra a pessoa que, observando que planos nunca correspondem adequadamente à realidade, sugere que não façamos plano algum. Além disso, se não soubermos quais as consequências de “cruzar os braços”, e não pudermos comparar esta opção às alternativas, a consequência lógica seria a “paralisia da decisão”, e não a adoção da *passividade* como solução *default*. Mas, se tal comparação for possível, então não precisamos sequer identificar a solução *ótima* para a decisão sob incerteza – basta mostrar que tende a ser melhor agir (ainda que sob algumas restrições) do que não agir.

Também não é simples colocar de lado a “racionalidade instrumental”. Embora haja várias razões para se defender a legitimidade de decisões democráticas, elas pouco têm a ver com a incerteza sobre o futuro; mesmo os representantes eleitos estão tentando fazer previsões e julgamentos acertados. Se previsões justificadas fossem algo impossível, decisões tomadas por uma *assembleia* seriam, nesse aspecto, tão problemáticas quanto as tomadas por um especialista¹³³.

Devemos encarar com cuidado, porém, a crítica vulgar de que a Economia tem, cada vez mais, se tornado um “ramo da matemática”, ou que suas previsões seriam inúteis (por falhar demais), e ignorar seu caráter empírico. Embora ainda estejam longe do grau de precisão da fictícia Psico-História, da *Fundação*, de Isaac Asimov (livro que inspirou ao menos dois ganhadores do prêmio Nobel em Economia, Paul Krugman e Roger Myerson), a aspiração de fornecer previsões acuradas e empiricamente testáveis, baseadas em modelos matematicamente rigorosos, é um dos motivos pelo qual economistas têm feito

¹³³ Porém note que, de acordo com o teorema do júri de Condorcet (1785), numa decisão binária coletiva por voto majoritário simples, em que cada votante tem alguma probabilidade maior que 50% de tomar a decisão correta, a probabilidade de que a decisão final seja correta *augmenta* à medida que aumentamos o número de votantes. Afinal, cada votante adicional tem uma probabilidade maior que 50% de influenciar o resultado final em direção à melhor resposta.

progressos científicos notáveis – frequentemente aproveitando insights de outras disciplinas¹³⁴. Mesmo diante do risco de “cisnes negros”, economistas têm se saído bem em explicar fenômenos sociais e fazer previsões modestas. Como veremos a seguir, um dos temas em que tal progresso pode ser observado é na explicação de crises financeiras.

6. Racionalidade econômica

Um dos tópicos de intensa colaboração entre economistas e filósofos é a teoria do agente racional. A concepção “fraca” (no sentido lógico), subjetivista e instrumental de razão como *coerência entre meios e fins* tem como padrão a teoria da utilidade esperada subjetiva. Ela se aplica a *problemas de decisão* caracterizados por: (a) um agente, (b) um conjunto de ações (mutuamente excludentes) que ele pode tomar, e (c) um conjunto de estados de coisas (mundos possíveis). A decisão envolve uma “regra de aprendizado”, que define de que forma o agente representa estados de coisas: (i) começa com uma distribuição de probabilidade *a priori* que atribui um valor correspondente a cada mundo possível; (ii) à medida que recebe nova informação, o agente atualiza sua distribuição de probabilidades mediante *condicionalização* (usando uma regra apropriada, como o teorema de Bayes) – o que resulta numa distribuição de probabilidade *a posteriori* (Bostrom, 2014, 10).

O agente também precisa de uma regra de decisão, que define como ele *valoriza* estados de coisas; assim, atribuímos-lhe uma função-utilidade – uma ordem *completa*, *contínua* e *transitiva* de preferências, que associa um número(utilidade) a cada estado de coisas, representando sua “desejabilidade”. Ao decidir, o agente seleciona a ação com maior utilidade esperada; para tanto, ele deve computar: (d) para cada mundo possível M e cada ação A , a respectiva probabilidade *a posteriori* p de atingir o mundo possível M condicionada à

¹³⁴ É frequente que contribuições técnicas inovadoras à economia e às finanças sejam feitas por matemáticos, físicos e engenheiros, que transferem técnicas de seus respectivos campos de atuação. Por outro lado, não menos destacáveis são as contribuições de outras ciências sociais, como psicologia (que forneceu a base da economia comportamental), filosofia e política. Isso pode ser confirmado em nossas sugestões de leitura.

escolha da ação A ; (e) para cada mundo possível M , o produto ($p \times u$) dessa probabilidade p e da respectiva utilidade u ; (f) para cada ação, a soma de cada um desses produtos (Bostrom, 2014, 11).

Essa teoria tem como objetivo explicar como os agentes fazem previsões em relação à forma como valorizam os estados de coisas - de modo que as probabilidades relativas atribuídas sejam uma função da inclinação de aceitar uma aposta sobre eles. Como vantagens, podemos destacar sua consistência dinâmica que impede que outro agente lucre fazendo *arbitragem* com as suas crenças (chamados *dutchbooks*). Além disso, "colocar um número" em um julgamento o incentiva a manter uma "precisão calibrada" (como fazem os super-previsores de Tetlock e Gardner, 2015), para representar sua própria incerteza. Enfim, o cálculo de probabilidade nos permite agregar informações sobre as opiniões de diferentes agentes - como argumentam os defensores da hipótese dos mercados eficientes.

Contudo, essa teoria leva a uma "explosão combinatória": ninguém com capacidade de processamento limitada conseguiria conceber todos os estados de coisas \times ações possíveis. Pior ainda em economia, onde o raciocínio recursivo (você modela outro agente modelando você) leva a situações *computacionalmente intratáveis*, onde não podemos encontrar uma solução em tempo hábil (Christian & Griffiths, 2018, p 165). Essas limitações levam a um comportamento "satisficente" (*satisficing*), e não estritamente *otimizador* (Pedroso, 2014) e, frequentemente, implicam vieses e potenciais falhas (Kahneman, 2012). Por fim, a teoria subjetivista não seria sequer uma aspiração normativa adequada, pois leva a uma *sub-determinação* da racionalidade: não exclui atitudes extremas nem arbitrárias (como alguém que prefere ver o mundo acabar a machucar o dedo, ou que não se importa com o próprio sofrimento nas terças-feiras – Parfit, 2013).

7. Moeda, legitimidade e justiça

7.1. O que é dinheiro?

Quando pensamos em dinheiro, podemos evocar *Tio Patinhas* nadando entre moedas de ouro, em sua caixa-forte – como o dragão *Smaug*, de Tolkien, em seu covil na Montanha Solitária. Essa associação entre dinheiro e metais

está impregnada em nossa cultura: Marx o chama de “vil metal”, e Locke defendeu que, pelo direito natural, a libra esterlina seria, *essencialmente*, uma quantidade de prata (Appleby, 1976); trata-se da *teoria metalista* – a ideia de que o valor da moeda é definido por uma *commodity* especial. Até o Acordo de Bretton Woods (1944), vigia o *padrão ouro*, e os governos garantiam a *convertibilidade* de suas moedas– i.e., a possibilidade de trocá-las pelo valor correspondente em metal. Ainda há ecos dessa ideia na teoria por trás de criptomoedas – *commodities* eletrônicas que precisam ser *mineiradas*.

Mas dinheiro é mais do que isso. A moeda é uma *tecnologia social* que executa quatro funções:

- a) *meio de troca*: dispensando a “dupla coincidência de desejos” – você não precisa perguntar ao barista o que ele precisa em troca de um café, você pode pagar o **preço**.
- b) *unidade de conta / medida*: o dinheiro fornece uma medida comum que permite comparar produtos totalmente diferentes a partir de seus preços;
- c) *reserva de valor*: você pode guardar dinheiro para usá-lo no futuro (ao contrário de bens perecíveis);
- d) *padrão de pagamento diferido*: graças às outras três funções, você pode acumular *dívida* em dinheiro (i.e., adquirir um produto hoje para pagar amanhã).

Quando funciona adequadamente, o sistema monetário de uma sociedade é como um “balanço distribuído”, que registra o quanto cada pessoa poderia consumir em bens e serviços, de acordo com o quanto de dinheiro ela tem. Ao longo da história, diferentes bens e costumes foram usados para essas funções, refletindo diferentes concepções sobre a moeda. A teoria atualmente dominante é a da moeda como *crédito*: o que ora chamamos de dinheiro é, principalmente, a moeda fiduciária – o papel-moeda que temos em nossas carteiras, baseado na confiança do governo que o emite.

Note que, como tecnologias, nem todas as formas de moeda funcionam igualmente bem nos mesmos contextos; elas podem ter problemas peculiares de acordo com a respectiva sociedade. Um problema com o metalismo é que

costuma levar à **deflação**: dada a oferta limitada de dinheiro (ou da *commodity* vinculada), as pessoas podem privilegiar a função *reserva de valor* no lugar de *meio de troca* – i.e., guardam a moeda ao invés de usá-lo; isso tende a agravar recessões, na medida em que se refreia o consumo. Na verdade, as pessoas podem simplesmente deixar de usar a moeda deflacionária (que se converte numa forma de poupança) em favor de outro meio de troca – a chamada “Lei de Gresham”, segundo a qual “dinheiro 'ruim' toma o lugar do dinheiro 'bom'”; i.e., as pessoas vão preferir quitar dívidas e fazer pagamentos com a moeda de menor valor e usar a de maior valor como reserva. Com uma moeda fiduciária, o emissor pode se comprometer a manter um fluxo constante de moeda – uma diminuta expansão monetária, para acompanhar a expansão da economia real.

Então por que o metalismo foi tão popular – e ainda apela a tantas mentes? A resposta mais simples é que ele impõe uma limitação ao poder do emissor (em geral, o soberano) – ao passo que a moeda-crédito implica dar poder ao emissor, que pode financiar como quiser suas atividades, *criando* mais moeda e gerando **inflação** (i.e., um aumento geral de preços nominais, decorrente da perda de valor da moeda). Criar moeda não *gera* riqueza, apenas a redistribui (Piketty, 2014, p. 669); p. ex., um estado moderno poderia criar moeda para adquirir empresas ou pagar dívidas, diminuindo o valor real de seu passivo em moeda nacional. Os usuários da moeda teriam, então, de arcar com preços mais altos, ou renegociar seus salários para acompanhar a inflação – o que gera uma espiral inflacionária: todos buscam continuamente aumentar o preço de seus próprios serviços. Isso implica, na prática, transferir riqueza dos usuários da moeda para o emissor, de maneira análoga a um tributo¹³⁵ (a *seignorage*). Por isso, é comum que economias modernas limitem a capacidade do governo central de emitir moeda, transferindo essa função a um órgão relativamente independente e apartado do poder político, o Banco Central.

7.2. Como fazer dinheiro

Ocorre que, como muitas pessoas descobriram com a Crise de 2008,

¹³⁵ Na Europa Medieval, senhores feudais que cunhavam moedas metálicas cobravam uma taxa, a *seignorage*, cada vez que cunhavam novas moedas, com uma proporção diferente de metais.

em sistemas de reservas fracionárias e afins, instituições financeiras também são emissores *privados* de moeda – graças à contabilidade. Imagine que você deposita \$100 em um banco, que são registrados no sistema da instituição como um passivo na conta de depósitos. Ao invés de deixar os \$100 guardados numa gaveta, o banco me oferece um empréstimo de \$100, que eu aceito, e ele agora registra um ativo de \$100 na conta de empréstimos; mas, ao invés de sacar os \$100 em notas, eu os deixo em minha própria conta de depósitos, que posso movimentar *online*. Note que, agora, nesse micro “sistema”, há \$200 disponíveis em moeda: os \$100 que você tem em sua conta, e os \$100 que tenho na minha – que o banco *emitiu*; ou seja, ele multiplicou seus \$100 iniciais, e ainda tem \$100 em seu cofre, que pode usar para oferecer crédito a outrem. Tudo isso funciona muito bem – desde que não tentemos sacar nosso dinheiro ao mesmo tempo.

Para influenciar a oferta de moeda, Bancos Centrais atuam sobre o crédito bancário; uma das alternativas é limitar a multiplicação de moeda reduzindo a capacidade de os bancos usarem depósitos para conceder crédito – p. ex., obrigando-os a manter uma parcela dos depósitos intocada (os depósitos compulsórios), ou a reduzir sua “alavancagem” (i.e., exigindo que aumentem a proporção entre o capital social e seus ativos). Principalmente, quando o Banco Central aumenta a taxa básica de juros da dívida pública, ele está desincentivando os bancos de conceder crédito a outros agentes; quando diminui a taxa de juros, os bancos aumentam a oferta de crédito – e, supondo que buscam maximizar seu retorno esperado, eles privilegiariam os investimentos mais eficientes¹³⁶.

Historicamente, a frequência de crises bancárias se reduziu com a consolidação dos respectivos bancos centrais – nos Estados Unidos, após a consolidação do sistema do *Federal Reserve* (e a regulação restritiva da Lei Glass-Steagall), durante o *New Deal*; e, no Reino Unido, após o pânico de 1866

¹³⁶ Ou não: mesmo que um banco faça empréstimos que ficarão inadimplentes no longo prazo, no curto prazo pode *parecer* um negócio lucrativo; trata-se de um exemplo de **risco moral** e de **conflito intertemporal** – que discutiremos melhor na seção seguinte. Mas não se assuste: seu dinheiro provavelmente está mais seguro num banco que debaixo do colchão; além da regulação e da supervisão financeira, a maioria dos países (inclusive o Brasil, por meio do FGC) fornece seguro para depositantes.

e a reforma do Banco da Inglaterra. Mas isso não impediu a crise de 2007-2008, causada pelo estouro da bolha imobiliária americana de 2006. Ela teria sido causada pela expansão monetária a partir de 2000 (justificada para contrapor os efeitos negativos sobre a economia da “bolha do *dot-com*” e do Onze de Setembro), ao financiamento barato por um amplo mercado não regulado (o *shadow banking*) de títulos lastreados em créditos imobiliários. Além disso, a inovação em finanças teria levado a operações cada vez mais complexas, cujos riscos se tornaram difíceis de avaliar (MacMillan, 2014).

Os governos, junto com os Bancos Centrais, agiram de forma inédita para resgatar empresas e bancos; apenas nos EUA, o *bail-out* haveria custado US\$ 700 bilhões. Isso foi duramente criticado: usar recursos dos contribuintes para salvar bancos e investidores (que voluntariamente assumiram certos riscos) violaria valores como igualdade e responsabilidade – mormente porque os devedores pobres não foram socorridos (Carter, 2020, p. 483). O problema é que, graças à alta conectividade do mercado financeiro (em especial, o mercado interbancário), a quebra de *um bancosistemicamente relevante* impacta a capacidade de outros agentes financeiros quitarem seus débitos, levando a uma *corrida bancária*: quando surge a notícia de que um banco está sob risco, os investidores correm para liquidar seus investimentos nos demais bancos, com medo de que eles também sejam afetados – o que resulta numa *profecia auto-confirmatória*, eliminando a liquidez dessas instituições (Peres, 2020). Ao invés do concurso de beleza de Keynes, tem-se uma *race to the bottom*¹³⁷.

A cada nova crise financeira, a influência dos Bancos Centrais aumenta, e se retoma o debate sobre sua autonomia; lembre-se que não votamos nos

¹³⁷Num exemplo de *race to the bottom*, os agentes devem dizer um número inteiro dentro dum intervalo finito; o vencedor é aquele que enunciar o número mais próximo do antecedente à média dos demais “lances” (arredondada para baixo) – o “resultado” do jogo. P. ex., se eu disser “10”, você disser “20” e um terceiro disser “15”, ele ganha. Agentes perfeitamente racionais sempre convergiriam para o menor número do intervalo; todavia, em situações reais de *race to the bottom*, os resultados nunca coincidem com tal solução. Ainda mais se a recompensa ao vencedor for uma função do resultado (i.e., o número dito pelo vencedor): suponha, p. ex., que o intervalo é $[0, 100]$ e que a recompensa prometida seria o dobro do resultado, em dinheiro, dividido entre os vencedores; nesse caso, o raciocínio do agente racional implicaria a resposta “0” – o que implica premiar o vencedor com nada!

diretores dos Bancos Centrais – o objetivo é justamente deixá-los distantes da política, para que possam exercer sua complexa missão macroeconômica. A maioria de nós sequer vislumbra o complexo sistema que subjaz ao nosso uso cotidiano do dinheiro; como poderíamos influenciar de forma eficaz essas questões técnicas? Ainda, pressionar autoridades monetárias por meio de representantes políticos costuma gerar desconfiança dos investidores e instabilidade econômica no médio prazo.

Isso acarreta um problema de confiança – análogo à discussão sobre os poderes dos juízes na interpretação da lei: os cidadãos, em nome e no interesse de quem essa autoridade age, podem se ressentir de quão pouco são representados por ela (que não elegeram, e que pouco parece responder a suas expectativas). E assim como um sistema jurídico pode ser visto como uma hierarquia de instituições e normas, um sistema financeiro pode ser visto como uma hierarquia de instituições e sujeitos, mediada por relações econômicas – e, nos dois casos, pode-se perguntar se os resultados e a respectiva divisão de poder são adequados, equitativos ou justos (Pistor, 2017, p. 185-86).

7.3. Inflação, desemprego e desenvolvimento

Uma questão historicamente polêmica é a da correlação negativa entre a taxa de inflação e a taxa de desemprego – i.e., um governo precisaria escolher entre diminuir o desemprego (estimulando a demanda agregada) ou a inflação (reduzindo a oferta de moeda – principalmente por aumento da taxa de juros). Economistas se dividiram, então, sobre a importância de cada um dos lados desse dilema: primeiro, a inflação equivale a uma taxa imposta sobre toda a sociedade, e pode impactar especialmente as pessoas que não dispõem de educação financeira; segundo, o desemprego implica o sofrimento direto de pessoas que perdem seus meios de sobrevivência.

A partir dos anos 70, a escola monetarista de Milton Friedman passa a defender que esse *trade-off* é mais complexo: existiria uma “taxa natural” de desemprego – e buscar reduzi-la, por meio da política monetária, seria ineficaz no longo-prazo. Isso nos desloca de uma discussão política para uma análise propriamente macroeconômica. Atualmente, essa teoria é atacada pela Teoria Monetária Moderna, defendida por André Lara Resende, no Brasil, segundo a qual um país não precisaria se preocupar com déficit em sua própria moeda.

Essa teoria é muito criticada – as condições em que um país poderia aplicá-la com sucesso (se existem) seriam raras, e acarretariam o risco de, diante de um choque externo, levar à hiperinflação ou a uma crise da dívida pública, como tantas vezes observado no século XX.

Embora essa seja uma discussão macroeconômica majoritariamente empírica, ela tangencia importantes questões políticas sobre liberdade, igualdade e solidariedade. Até que ponto o governo deveria intervir na economia com o objetivo de influenciar as decisões dos agentes e gerar empregos? Se existe um *trade-off* entre inflação e desemprego, quais os níveis socialmente aceitáveis? Mas, principalmente, não seria melhor proteger não o emprego, mas o *trabalhador* – i.e., expandir a seguridade social e o seguro-desemprego?

Tal debate se torna mais urgente quando se trata de propostas de implementação de *renda básica universal*. Uma delas seria uma renda básica como *instrumento de política monetária* (Mcmillan, 2018, p. 174-175): ao invés de agir sobre a oferta de crédito, o Banco Central poderia diretamente aumentar o saldo na conta bancária de cada cidadão. De fato, um dos motivos por que governos forneceram renda aos trabalhadores afetados pela pandemia de Covid-19 foi mitigar o efeito deflacionário da crise de 2020; o problema desse instrumento é que seria necessariamente instável – teria de variar de acordo com a necessidade de moeda na economia, não de acordo com a necessidade de recursos dos cidadãos.

Por isso a versão mais popular da proposta é a de uma garantia de renda de cidadania – enfatizada pela previsão de substituição de mão-de-obra pela automatização. Mesmo que, no longo prazo, o crescimento econômico proporcionado pelo aumento de produtividade *tenda* a gerar novos empregos, é preciso notar que pode ser difícil readaptar trabalhadores especializados (p. ex., datilógrafos); e as pessoas continuam precisando de renda enquanto os novos bons empregos não surgem. Nesse ponto, a automatização poderia aumentar os efeitos socialmente negativos da desigualdade social. Afinal, no longo prazo, ou estaremos todos mortos, ou nossos empregos serão radicalmente diferentes – e altamente automatizados. Essa não é uma preocupação tão recente: Marx já havia apontado essa possibilidade no seu *Fragmento sobre máquinas* (Spence, 2019). Se isso parece ficção científica para o leitor, peço que recorde nossos comentários sobre o progresso do século XX (que soariam como ficção para qualquer um nascido em 1900) – ou que busque interagir com Processadores de

Linguagem Natural de última geração (como, no momento em que escrevo, o GPT-3, lançado pela *OpenAI*). É recorde que não é preciso que automatizemos *todo* o trabalho humano – basta que haja ganhos de produtividade grandes e súbitos: se a produtividade de uma empresa dobrar sem que haja um aumento da demanda, ela poderá funcionar com metade de seus empregados.

Em *Possibilidades Econômicas para Nossos Netos*, de 1930, Keynes (1963) vaticinou que, graças ao aumento da produtividade e à redução da desigualdade social, a semana laboral passaria a ter apenas 15 horas, e o resto do tempo seria dedicado ao lazer; ele errou nesse ponto, mas é interessante notar o porquê – a começar pela demografia (tópico que economistas às vezes negligenciam). Ele, de fato, acertou quanto ao aumento da produtividade; mas note que você vive numa cultura muito mais rica, com arte e conhecimento ao alcance de sua mão, e provavelmente tem uma vida materialmente mais confortável do que Lord Keynes teve no ápice de sua carreira: mais saúde, segurança alimentar, transporte, eletrônicos, recursos computacionais, possibilidades de lazer... Além disso, o trabalho de boa parte das pessoas deixou de ser manual e repetitivo, embora possa eventualmente ser entediante; e os rendimentos acumulados do trabalho não são imediatamente consumidos, mas poupados para sustentar longas aposentadorias e filhos – que, cada vez mais, só entram no mercado de trabalho após a conclusão de um longo período de estudos. Para tornar possível essa ampla oferta de serviços e bens, ainda trabalhamos mais do que 15h/semana.

Keynes *subestimou nossas perspectivas e necessidades*. Não cometamos o mesmo erro.

8. Ética nos negócios e responsabilidade social

Além de grande expoente da escola monetarista, Thomas Friedman (2014) também é conhecido por sua doutrina sobre a responsabilidade social da empresa – que se reduziria a gerar lucros para os sócios. Claro, as empresas deveriam obedecer à legislação (o que preclui, por exemplo, corromper autoridades), e até mesmo a algumas proibições éticas básicas (como a exploração de trabalho escravo); mas, fora isso, numa economia de mercado, a firma serve para maximizar o lucro do sócio. Essa conclusão pode ser defendida pelo argumento:

- a) do **dever fiduciário** dos gestores de agir no interesse dos sócios – i. e., a firma é uma ficção jurídica, e seu patrimônio pertence, em última instância, aos sócios. Esse dever visa a mitigar o problema do **agente-principal**: como confiar que os gestores não estão atuando em prol de seus interesses pessoais, ao invés do dos sócios?
- b) da **disciplina de mercado** – i. e., a firma precisa ser eficiente e maximizar o retorno sobre o investimento, ou, num mercado competitivo, perderá sua posição para seus concorrentes¹³⁸.

Essa discussão se tornou o cerne da ética nos negócios. Para a doutrina Friedman, existe uma desanalogia entre o papel do indivíduo como agente moral e o da empresa. Note como essa é uma questão sobre ética – sendo, em princípio, logicamente independente de teses sobre o papel do estado em economia política: você pode ser liberal e defender que as empresas têm obrigações morais para com a comunidade, mesmo sem qualquer dever legal definido a respeito, ou, pelo contrário, defender que cabe ao governo, não às empresas, submeter-se ao escrutínio moral, mesmo que você seja socialista. Subjaz a esta teoria um argumento de divisão moral do trabalho, segundo o qual a responsabilidade social e ambiental deve ser tarefa dos governantes, ou talvez da sociedade civil, enquanto o setor financeiro deve se concentrar em aumentar e manter o capital (Sandberg, 2016).

É raro que teóricos e empresários abertamente subscrevam à doutrina Friedman; ironicamente, fazê-lo poderia prejudicar a *reputação* da firma e levar a perdas econômicas. Ao invés disso, é comum que adotem algo como a doutrina do relatório *Freshfields*, segundo o qual o dever fiduciário seria compatível com a existência de muitos tipos de situações de investimento, em que gestores

¹³⁸ Isso não quer dizer que a firma *explicitamente* busca maximizar seus lucros, ao contrário do que prega a doutrina Friedman; num mercado competitivo, as empresas *sobreviventes* serão aquelas cujo comportamento adaptativo pode ser descrito como maximizando o retorno sobre o investimento – da mesma forma que plantas não visam a maximizar a transmissão de sua carga genética, mas, graças à seleção natural, se comportam como se o fizessem (Alchian, 1950). Isso é relevante porque, num ambiente de negócios onde normas de responsabilidade social são observadas, as firmas podem ser descritas como adotando tal comportamento adaptativo.

deveriam considerar os interesses de outras partes interessadas, como empregados, fornecedores e comunidades locais. Então, pode ser permitido, ou mesmo obrigatório, considerar como as empresas se comportam do ponto de vista social ou ambiental, quando isso: for do interesse “ético” *consensual* dos sócios, ou não prejudicar a expectativa de lucro (i.e., o interesse social pode ser um argumento de *desempate* da decisão negocial), ou quando afetar o desempenho de longo prazo da firma (p. ex., se impactar o preço futuro das ações, ou acarretar um risco de responsabilização judicial). Além disso, é plausível que uma cultura de responsabilidade social seja correlacionada com a observância de obrigações legais e morais básicas (Lo, 2015); i.e., se você quer uma firma que apenas obedeça à legislação, é preciso que ela aspire a muito mais do que gerar retorno para o sócio.

A própria ideia de que os sócios (principalmente em companhias abertas, onde a participação social é um título de crédito com liquidez) são proprietários da empresa, e não um tipo muito especial de investidor, é questionável, e a descrição legal da relação entre a firma e o sócio pode variar muito de acordo com a legislação do país. Note que há uma assimetria entre a perspectiva de perdas e ganhos do sócio, o que pode levar a um **risco moral**: como a perda possível é limitada ao capital social (ou o valor de compra da ação), mas não há limites para a possibilidade de lucro, pode ser do interesse do sócio que a firma opere de maneira arriscada e alavancada – transferindo risco para os demais credores e para a sociedade em geral. Se um banco alavancado falir porque emprestou dinheiro demais para poluidores, seus investidores, empregados e as vítimas da poluição ficarão sem ressarcimento; mas os sócios perderão “apenas” o capital aplicado. Neste caso, não deveria haver um princípio para limitar essa assimetria¹³⁹?

¹³⁹ Na verdade, o exemplo é ruim: no Brasil, há exceções legais para a limitação de responsabilidade do sócio, além de regimes especiais de responsabilidades para administradores e controladores em alguns setores regulados, como bancos; no exemplo, além de ser punível por gestão temerária, um banqueiro poderia ter que responder pelas dívidas da instituição com seu próprio patrimônio.

9. Reputação e moralismo

Ativistas frequentemente são bem-sucedidos ao impor custos reputacionais a empresas, obrigando-as a, p. ex., eliminar a utilização de ovos de frangos criados em gaiolas – uma forma particularmente cruel de criar aves. Além disso, alguns movimentos ativistas, como a PETA (*People for the Ethical Treatment of Animals*), tem adotado como estratégia comprar ações dessas empresas, a fim de melhor fiscalizá-las e de apresentar suas reivindicações aos demais acionistas e executivos. Se a doutrina Friedman estiver correta, então esses ativistas deveriam ser repreendidos – eles estariam usando a firma para fins “escusos”; mas por que a empresa deveria servir apenas à maximização do lucro do sócio, e *jamaiz* a seus interesses éticos?

Uma alternativa é enfraquecer a doutrina Friedman e apelar para uma analogia com algumas vertentes de relativismo moral, para as quais não há princípios morais *verdadeiros* – há apenas algo como a “aparência” de um princípio: as pessoas agem de acordo com regras para sinalizar que estão aptas a cooperar em sociedade (e se beneficiar disso). Dessa forma, tais empresas não têm qualquer *razão moral* para ceder à pressão, apenas razões prudenciais: elas querem ser vistas como empresas “boazinhas”, com reputação melhor que as concorrentes.

Além disso, note que é possível haver a imposição de limites socialmente deletérios aos mercados, com base em intuições morais irrefletidas. Por exemplo, Sandel (2012) defende que restrições morais e legais se justificariam porque há atividades cujo valor é *diminuído* quando são objeto de transações econômicas; isso cria uma restrição econômica causada por *repugnância moral*, como no caso de mercados de órgãos (Roth, 2007), mesmo *que tais restrições resultem em menos vidas sendo salvas*. Por exemplo, há diversos países (no Brasil e na Europa) que proíbem a compra e venda de plasma sanguíneo; conseqüentemente, seus estoques de plasma são insuficientes, e precisam ser repostos mediante importação de países onde a venda é permitida – i.e., o governo proíbe que você pague pelo plasma dentro de sua jurisdição, mas então o *compra* de países onde a venda é liberada (Jaworski, 2020).

Uma das ideias mais importantes em economia é que, mais que por juízos morais, as pessoas são movidas por *incentivos* - normalmente ligados ao

interesse individual; os economistas, então, tentam identificar os esquemas de incentivos que possam ser implementados por uma instituição para atingir determinados fins. Um exemplo clássico: no séc. XVIII, o Império Britânico expatriava seus detentos para a Austrália; graças às condições monstruosas dos navios, cerca de um terço dos “passageiros” morria na viagem, e os demais chegavam feridos e famintos. Apesar da revolta da opinião pública e da crítica dos parlamentares, nada parecia resolver a situação; até que um economista sugeriu que, ao invés de pagar por cada prisioneiro embarcado o governo passasse a pagar por cada prisioneiro que chegasse vivo à Austrália. Quase que imediatamente, as mortes durante a travessia caíram 99% (Cowen & Tabarrok, 2013, p. 1)¹⁴⁰.

Isso **não** implica uma objeção geral contra a noção de responsabilidade social; porém, pelo contrário, estamos tentando mostrar como um papel importante dos economistas (e outros *designers* de instituições) é desenhar instituições que efetivamente satisfaçam nossas aspirações morais, ao invés de apenas *parecer* que as satisfazem.

Imagine que três empresas construíram uma reputação como socialmente responsáveis: (i) uma empresa que prepara um belo, mas inócua, relatório de responsabilidade social, enfatizando como, p. ex., economizou em papel com a digitalização de processos; (ii) uma empresa que observa as normas sociais de segregação no Sul dos Estados Unidos, na década de 50; (iii) uma empresa que, como a Merck (Moriarty, 2017), cria e distribui gratuitamente

¹⁴⁰ Note que não é óbvio descobrir quais os incentivos adequados para o problema que você quer resolver: no séc. XIX, paleontólogos europeus pagavam às populações nativas por pedaços de fósseis e ossos; consequentemente, quando encontravam ossos grandes (que teriam maior valor científico), elas os quebravam em vários pedaços menores para aumentar seus lucros (Mullainathan & Shafir, 2013, p. 15). Essa dependência de incentivos individuais pode ser um tanto lamentável – seria mais fácil que nos coordenássemos, se fôssemos menos egoístas. Porém, para primatas que evoluíram resolvendo alocações de recursos mediante violência, e que hoje costumam se sentir especiais porque compram um carro novo (ou sentem inveja porque outros o fazem), já é um grande progresso reconhecer esses aspectos de nossa natureza e usá-los para melhorar a vida uns dos outros. Ao invés de lamentar a escassez de boas intenções, pode ser mais prático dirigir nosso altruísmo para projetos que efetivamente funcionem.

um tratamento contra a oncocercose, prevenindo a cegueira de milhares de pessoas em países em desenvolvimento. Se a ideia de responsabilidade social é apenas uma estratégia para melhorar a reputação da empresa (e assim continuar a maximizar o lucro do sócio, indiretamente), as três empresas estão na mesma situação. Afinal, o que nos permite distinguir entre o caso de: (i) executivos estão sendo cínicos, ao se apresentarem como socialmente responsáveis, (ii) há uma restrição social indevida aos negócios, e (iii) existe uma justificativa moral sólida para que a empresa (e seus gestores) aja de certa forma¹⁴¹? Como podemos identificar maneiras *corretas* de formar uma reputação de responsabilidade social – i.e., os casos onde a reputação é devidamente obtida? A única forma de fazer tal distinção é levando a sério a deliberação sobre a existência de princípios morais subjacentes; nisso, as pessoas agem como se de fato acreditassem que estão discutindo um princípio moral (quer estejam certas, quer erradas). Se tudo que há é “aparência e reputação”, essa discussão não faz sentido. Mesmo que, em última instância, a resposta para essa questão seja uma concepção *convencionalista* de ética nos negócios, seria preciso identificar quais os motivos e princípios que orientam tais convenções – que permitem que mesmo o mais cínico dos executivos finja estar agindo de forma ética. Nisso, a analogia entre ética nos negócios e ética individual se mantém.

O argumento sobre a divisão de trabalho é um pouco mais forte: governos, e não mercados, deveriam ser o espaço para a deliberação moral. Friedman está correto em frisar a importância do lucro do sócio: num sistema de mercado funcionando em condições regulares (o paradigma dos livros de introdução à Economia), isso implica que a firma está fornecendo aos consumidores serviços que eles desejam, remunerando seus trabalhadores de uma maneira que eles aceitam, e observando restrições regulatórias adequadas. Porém, o fato de que algumas obrigações morais devem ser determinadas pela legislação não justifica concluir que *todas* devam sê-lo: a definição da tarefa dos governos não exclui todas as razões morais dos indivíduos para fazer o bem, principalmente se a atuação dos governos for insuficiente. Afinal, o fato de alguém ter sido designado para uma tarefa não implica que ninguém mais tenha uma razão moral para executá-la. Mesmo que você não seja um socorrista, você

¹⁴¹ Mesmo que os gestores e acionistas não sejam motivados por essa justificativa – basta que compreendam que é uma justificativa reconhecida por outras pessoas.

tem uma razão moral para prestar auxílio ao observar um acidente – e essa razão é ainda *mais forte* se ninguém mais o fizer.

Alguém poderia afirmar que isso implica que os *indivíduos* deveriam intervir, mas não as empresas; no entanto, os mesmos motivos que explicam por que uma empresa é economicamente mais eficaz do que um indivíduo para fornecer bens e serviços também explicam por que, em algumas situações, ela pode ser mais eficaz para fazer o bem: economias de escala, divisão de trabalho por especialização, poder de barganha, transparência, menor aversão ao risco, etc. Nesse caso, o argumento sobre a divisão de trabalho pode se voltar *contra* o defensor da doutrina Friedman: se uma empresa é o agente mais bem posicionado para praticar uma ação moralmente boa, isso parece ser uma razão moral relevante para o fazer – ao menos se não houver uma razão contrária.

10. A Tragédia do Horizonte

Um importante aspecto da discussão sobre responsabilidade social é o conflito intertemporal entre os interesses dos sócios: os interesses de maximização do lucro dos sócios de hoje podem ser contemplados em detrimento dos interesses dos sócios futuros (se os houver). O problema é que o escopo do planejamento de um governo ou uma empresa raramente ultrapassa cinco anos; mesmo a política monetária, em geral, mira dez anos à frente. Isso leva ao que Mark Carney (2015), então Presidente do Banco da Inglaterra, chamou de a *Tragédia do Horizonte*: negligenciamos projetos e problemas de longo prazo – em especial os necessários para responder ao aquecimento global. Desde então, autoridades e instituições financeiras ao redor do mundo têm se reunido em torno da *Network for Greening the Financial System*, a fim de coordenar ações para mitigar riscos financeiros decorrentes das mudanças climáticas, os “cisnes verdes” – como destacado por recente relatório do *Bank for International Settlements* (Bolton et al., 2020).

Um dos desafios é que costumamos “descontar” o valor do futuro. Imagine que eu lhe ofereça (a) \$100 hoje, ou (b) uma quantia *nem* um ano; quanto deveria ser *n* para que você preferisse a segunda opção? Esse valor adicional de *n* captura sua *taxa de desconto*. Filósofos e economistas concordam que é aceitável ter uma taxa de desconto positiva, para refletir a incerteza do futuro (p. ex., a possibilidade de morte) e o custo de oportunidade (você pode

investir n hoje ter ainda mais no futuro); mas alguns economistas, a partir das preferências das pessoas, também concluem que há uma taxa de desconto *intrínseca* – i.e., que o presente, por si só, é mais valorizado que o futuro, independentemente da incerteza.

Governos e instituições usam uma *taxa de desconto social* para analisar o valor de projetos de longo prazo; alguns defendem que ela deve ser neutra em relação ao futuro (Stern, 2007): devemos tratar igualmente os interesses das pessoas, em qualquer momento do tempo, e apenas descontar o futuro de acordo com a incerteza (p. ex., com o risco de que alguma catástrofe global ocorra). Outros defendem uma taxa mais alta, semelhante ao que as pessoas preferem em suas escolhas individuais; mas há uma importante desanalogia entre o cenário pessoal e o coletivo: enquanto nossas perspectivas pessoais e nossa expectativa de vida são altamente limitadas, o mesmo não vale para sociedades amplas – não há um limite natural, uma “expectativa de vida” para a espécie humana.

Outros, porém, aduzem que devemos ter uma alta taxa de desconto para o futuro próximo, e uma baixa para o futuro mais distante (Nordhaus, 2007); isso reflete como as pessoas frequentemente agem – com **desconto hiperbólico**. Diante da escolha entre (i) \$10 agora e \$11 em uma semana, muitas pessoas preferem a primeira opção; mas com a escolha entre (ii) \$10 em um ano ou \$11 em 372 dias, elas preferem a segunda. Embora isso não seja uma contradição em sentido estrito, resulta em *planejamento inconsistente*, ou *inconsistência temporal*: se em (ii) repetirmos a oportunidade de escolha um ano depois, a mesma pessoa irá revertê-la, e preferir o dinheiro imediatamente! Isso é um fato cotidiano da psicologia humana: estamos acostumados a saber que podemos nos arrepender de decisões, por preguiça ou intemperança.

Isso pode tornar problemas globais de longo prazo particularmente difíceis de resolver, ainda mais do que problemas usuais de bens públicos, pois nenhuma geração tem incentivos adequados para adotar medidas em benefício de seus sucessores. Alguns economistas sugerem resolver o problema do efeito estufa internalizando os custos sociais de emissão de carbono num mercado de direitos de emissão negociáveis (Tirole, 2020); no entanto, dado o desconto hiperbólico, é razoável ser cético quanto à capacidade de os preços dos “créditos de carbono”, por si só, efetivamente refletirem os custos futuros correspondentes.

Voltando à analogia com o cenário individual: geralmente resolvemos inconsistências temporais, adotando medidas ou assumindo compromissos, no presente, que *restringam* as nossas opções no futuro – como Ulisses, que, na *Odisseia*, se amarra ao mastro de seu navio para evitar ser levado pelo canto das Sereias. Da mesma forma, se você tem receio de sucumbir à gula, uma alternativa é não ter guloseimas em casa; e se você tem receio de mudar de ideia no exemplo (ii), pode prometer a um amigo pagar uma multa por preferir a primeira opção.

É difícil extrapolar dessas soluções para o cenário intergeracional: como podemos fazer acordos com pessoas que só existirão quando estivermos mortos? Uma solução criativa seria “pedir emprestado ao futuro” (i.e., dívida de longo prazo): John Broome e Duncan Foley (2015) sugerem criar um Banco Climático Mundial, com o orçamento proveniente de taxas sobre emissões de carbono, para emitir títulos de dívida de longo prazo que financiariam a transição para energias renováveis (e compensariam as pessoas prejudicadas no presente). Não seria a solução mais *justa*: as gerações futuras, que serão obrigadas a pagar pela transição verde, ainda não lançaram carbono algum na atmosfera; mas é uma solução *compatível com os incentivos* que temos, e muito melhor do que não fazer nada – elas pagarão alegremente por evitarem uma catástrofe climática¹⁴². Outra dúvida é como isso poderia ser financeiramente sustentável, dado o alto nível global de endividamento público e privado.

Na prática, porém, nenhuma dessas iniciativas teria qualquer chance de

¹⁴² Outra ideia criativa seria usar desconto hiperbólico de investidores para benefícios sociais de longo prazo; p. ex., na proposta *Windfall Clause*, cada grande corporação assumiria o compromisso irrevogável de, caso descubra uma tecnologia revolucionária que lhe permita concentrar uma parcela relevante do PIB mundial (como inteligência artificial geral) nos próximos 50 anos, distribuir uma fração dos lucros (p. ex., 10%) com o restante da humanidade (O’Keefe et al., 2020). É como prometer a um amigo que, se ganhar na loteria, você doará parte do prêmio: individualmente, para cada empresa, o ganho reputacional de tal compromisso é relevante, e o custo presente é pequeno – já que a probabilidade *ex ante* de efetuar tal pagamento é pequena, e mesmo se tivesse de efetuar-lo, não abalaria sua posição futura no mercado. Do ponto de vista coletivo, a cláusula também é uma vantagem, pois funciona como um seguro contra o cenário em que uma única firma viesse a deter posição dominante sobre os recursos econômicos globais.

ser implementada com sucesso sem o apoio de empresas, governos e sociedade civil. Talvez o principal argumento contra a doutrina Friedman seja simplesmente o reconhecimento de que, sem uma concepção de ética nos negócios que permita a companhias relevantes assumirem compromissos duradouros com as comunidades (incluindo a comunidade global e as gerações futuras), não apenas essas companhias, mas nossas próprias sociedades serão insustentáveis no *longo* prazo.

11. Ética em finanças – responsabilidade social para fundos de pensão

No Brasil, os bancos são obrigados a ter suas próprias políticas de responsabilidade socioambiental, graças uma Resolução do Conselho Monetário Nacional. A justificativa legal é a mediação entre os interesses de longo e de curto prazo dos sócios: essas políticas visam a evitar futuras perdas decorrentes de danos socioambientais. Uma justificativa mais ampla, porém, é que o setor financeiro deve assumir tal responsabilidade porque desempenha um papel central na alocação de capital e no provimento de liquidez a outras atividades; assim, suas ações socioambientais têm impacto ampliado sobre o restante da economia – p. ex., ao exigir de seus clientes a observância da legislação ambiental, os bancos podem ter um efeito análogo ao de órgãos de fiscalização.

Pelos mesmos motivos, seria útil se uma norma similar obrigasse a adoção de políticas de responsabilidade socioambiental por parte de outro tipo de empresa financeira com larga influência sobre a economia, os *fundos de pensão*, onde trabalhadores investem suas economias para se aposentar. Segundo a OECD (Ervin, 2008), estes fundos detêm, em média, ativos equivalentes a 77% do PIB de seus países correspondentes; só o *Government Pension Fund Global* da Noruega (que recentemente anunciou seu desinvestimento em multinacionais brasileiras que não observavam sua política de responsabilidade socioambiental), tem cerca de U\$ 1 trilhão em ativos, incluindo 1,4% do mercado global de ações. No Brasil, os fundos de pensão de funcionários de empresas estatais (Funcef, Petros, Previ) detêm participações relevantes em empresas nacionais (Musachio & Lazzarini, 2015, 19). Alguns fundos fechados de fato mencionam o termo “responsabilidade socioambiental”, mas, atualmente, eles não têm qualquer obrigação específica a respeito.

Portanto, se adotássemos um princípio que exigisse alocar responsabilidade onde ela pode ser desempenhada de forma mais eficiente, isso deveria incluir fundos de pensão. Trata-se dos veículos de investimento com maior perspectiva de longo prazo – sendo, pois, uma forma conveniente de mitigar o problema da Tragédia do Horizonte. Ainda, esses fundos não são empresas normais em mercados competitivos; trata-se de um setor altamente regulado (que ou recebe recursos públicos, ou é subsidiado mediante incentivos fiscais – como dedução do imposto sobre aportes), e suas cotas integram as economias de trabalhadores, não podendo ser adquiridas ou vendidas livremente – ao contrário das ações de companhias abertas. Portanto, os principais argumentos da doutrina Friedman **não** se aplicariam aqui:

- a) estes fundos não se submetem à mesma **disciplina de mercado**, pois eles não competem diretamente entre si, e os cotistas não podem comprar e vender cotas com a mesma facilidade com que compram e vendem ações, e não há a mesma pressão por retorno no curto prazo – pelo contrário, há ênfase em retornos sustentáveis;
- b) o problema do agente-principal é mitigado principalmente por regras regulatórias, que aplicam interpretações particulares do dever fiduciário.

Conclusão

Peço desculpas por lançar uma dose tão alta de informações sobre o leitor, com explicações bastante curtas; não espero que você domine plenamente esses temas, mas que se sinta intelectualmente instigado pelas questões *filosóficas* supra. Lembre-se que elas também são problemas reais: cada nova crise econômica traz uma dose de sofrimento aos que têm diminuídas suas perspectivas de vida, e, a todos nós, a lembrança de que o desenvolvimento econômico observado no último século não é uma *lei da natureza* – é uma exceção na história humana.

Ao invés de rever os pontos discutidos, gostaria de focar apenas um. Começamos este texto destacando as vantagens de viver num mundo onde você pode agradecer a milhares de pessoas por seu café. Por outro lado, se cada pessoa em cadeias de cooperação tem uma parcela de *mérito* pelo resultado, o

mesmo vale para a *responsabilidade*. Nossa cultura e psicologia moral começaram a despertar para o fato de que essa responsabilidade difusa conflita com nossas noções éticas intuitivas, nas quais uma pessoa *determinada* é considerada responsável por causar um dano *específico* a outrem – i.e., onde vítimas podem apontar culpados por danos mensuráveis.

Em concepções tradicionais de razão prática, costumam-se separar as decisões e preferências em dois domínios: há um domínio pessoal, onde você procura satisfazer algumas de suas preferências individuais, condicionadas a restrições morais; e outro domínio social ou político, onde você vota e participa de decisões (o que economistas chamam de *escolha social*) que visam a alguma concepção de bem estar coletivo, abrangendo outras pessoas com quem você coopera. Essa separação é importante e útil: assim como (provavelmente) ninguém conhece suas preferências e necessidades melhor do que você, é provável que você não conheça tão bem as de outras pessoas; ela nos permite, pois, demarcar uma esfera privada onde nossas decisões são “soberanas”, onde cada um pode acertar e errar por sua própria conta e risco, protegido de coerção externa.

Frequentemente, filósofos e economistas defendem que essa ideia de esfera privada é um componente característico dos mercados. Mas, quando analisamos mercados mais detidamente, a fronteira entre os dois domínios tende a colapsar: suas decisões acarretam externalidades e custos de oportunidades para outros além do que você pode prever individualmente. Cada unidade monetária de suas transações econômicas é não só uma sinalização de suas preferências e custos de oportunidade, mas também um “voto na função de escolha social”: o resultado do mercado é “apenas” um agregado de escolhas de indivíduos e firmas. Ainda que as consequências desse agregado sejam difíceis de antever, isso não o exime de lembrar que você faz parte disso; e mesmo que sua influência marginal seja pequena, lembre-se que isso vale igualmente para muitas outras decisões, e para a maioria das pessoas (p. ex., é absurdamente improvável que seu voto seja decisivo em qualquer eleição política). Conquanto mercados sejam muito bons em agregar informações relevantes aos preços de ativos e produtos, eles não podem *tomar* decisões morais por você; porém são uma importante lembrança de que, na prática, você não as toma solitariamente.

Questões de revisão

1. Qual a importância moral do desenvolvimento econômico?
2. Quais os problemas que o desenvolvimento econômico trouxe – que não existiam (ou eram mínimos) até o século XX?
3. Por que o “problema da incerteza” é importante para a economia?
4. Por que preços podem ser chamados de “sinais enrolados em incentivos”?
5. Por que defensores da HME criticam a intervenção estatal em mercados?
6. Qual o aspecto paradoxal da HME?
7. Explique como a noção de reflexividade afeta o problema da incerteza? Mais: associe essa explicação ao “concurso de beleza de Keynes” e à HME.
8. O que é o dinheiro, segundo a teoria da moeda como *commodity*?
9. Quais as funções do dinheiro?
10. O que é o dinheiro, segundo a teoria da moeda como crédito?
11. Qual o problema *político* da teoria da moeda como crédito?
12. Qual a justificativa para o resgate de bancos durante a grande crise?
13. Qual o problema político gerado por esse resgate?
14. Qual o problema do risco moral?
15. Por que Bancos Centrais costumam ser apartados da política? Qual a vantagem e a desvantagem disso?
16. Qual a relevância política do debate econômico sobre inflação e desemprego?
17. Que vantagens podem ser apresentadas para programas de transferência direta de recursos – como a renda básica?
18. Por que dissemos que Keynes “subestimou” nossas perspectivas?
19. O que é a doutrina Friedman da responsabilidade social corporativa?
20. Qual a relação entre o dever fiduciário e o problema do agente-principal? Explique os dois conceitos.
21. Qual o “argumento da divisão moral de trabalho”?
22. Imagine que você é um investidor; com que base você poderia reclamar da atuação das seguintes empresas?
 - (A) A qual adquire um caro processo de reciclagem de papel;
 - (B) A qual se orgulha de cooperar com governos (alguns dos quais não democráticos), inclusive provendo acesso a dados pessoais de seus clientes;

(C)A qual desenvolve medicamentos para doenças tropicais negligenciadas – apesar de ser menos lucrativo do que tratamentos para doenças crônicas mais prevalentes em países desenvolvidos.

23. O que é a *A Tragédia do Horizonte*?
24. O que é a *taxa de desconto social* e como isso afeta a *Tragédia do Horizonte*?
25. Dê um exemplo onde você manifestou desconto hiperbólico – se você for um agente perfeitamente racional e temporalmente consistente, pode inventar um exemplo.
26. Por que o desconto hiperbólico pode levar a um planejamento inconsistente?

Questões de discussão

1. O estudante Jay Williams foi ao Sudão, para trabalhar para a *Christian Solidarity International*, e pagou pela libertação de 4000 pessoas que haviam sido escravizadas pelas facções da guerra civil (Cowen & Tabarrok, 2013, p. 65). A questão aqui é se a oferta de escravos é elástica: ao comprar a liberdade pessoas escravizadas, Williams sinaliza um aumento da demanda por escravos, o que leva a um aumento de seus preços, e incentiva os mercadores de escravos a capturar mais pessoas; estimar o quão elástica é esse oferta demanda uma análise econômica, claro. Suponha que possamos determinar isso com uma certeza aceitável, e descobrimos que, para cada duas pessoas libertadas por Jay, os mercadores de escravos são incentivados a escravizar mais uma pessoa – uma razão de 2:1. Como você avaliaria a atitude de Jay, nessa hipótese? Qual a proporção N:1 que faria você mudar de ideia em relação à atitude de Jay?
2. Em 2020, o *Government Pension Fund Global* (GPF) da Noruega anunciou que liquidaria seus investimentos em companhias que não estivessem de acordo com sua política de responsabilidade social; isso incluiu firmas de países em desenvolvimento. De um lado, alguns investidores alegam que, pela Hipótese dos Mercados Eficientes (numa versão enfraquecida), a atuação do Fundo é inócua: afinal, se o GPF está vendendo, é porque alguém está comprando – o GPF estaria apenas

transferindo esses ativos para outros compradores (que ficarão felizes em comprá-los). Você concordaria com esse argumento? Quais seus pontos fortes e fracos? Por outro lado, investidores de países afetados reclamam pelo motivo inverso: o GPFM prejudicaria a capacidade de financiamento dessas empresas, comprometendo as perspectivas das nações em desenvolvimento. Você concorda com este argumento?

Sugestões de leitura

Na verdade, ainda há poucos materiais de Filosofia da Economia traduzidos para o português; as entradas da *Stanford Encyclopedia of Philosophy* sobre o tema são fortemente recomendadas, mas hoje elas *ainda* não estão traduzidas – embora estamos trabalhando na tradução e publicação de volumes da Série Investigação Filosófica sobre o tema¹⁴³. Uma exceção é o livro do filósofo Weatherall (2015), que traz uma interessante história sobre como o mercado financeiro passou a absorver teorias matemáticas complexas para lidar com a incerteza. No entanto, alguns livros escritos por autores com outras formações em ciências humanas e sociais trazem reflexões fundamentais para qualquer um interessado no tema do ponto de vista filosófico, como Jean Tirole (2020), Amartya Sen (2011), Alan Greenspan (2013), o psicólogo Daniel Kahneman (2012), o historiador Yuval Harari (2015), e mesmo o investidor (com formação filosófica, ex-aluno de Karl Popper) George Soros (2008). Macmillan (2019) contém uma explicação bastante didática sobre a evolução e o atual funcionamento dos sistemas financeiros, de sua relação com a crise de 2008, e sugestões de reforma.

Referências

ACEMOGLU, D.; ROBINSON, J. (2013). *Por que as Nações Fracassam*: as

¹⁴³ Veja mais sobre a Série Investigação Filosófica, que publica volumes de traduções autorizadas de verbetes da *Stanford Encyclopedia of Philosophy*, editada pelo Grupo Investigação Filosófica da Universidade Federal do Amapá, pelo Núcleo de Ensino e Pesquisa em Filosofia da Universidade Federal de Pelotas, tal como também pela sua Editora, aqui: <https://wp.ufpel.edu.br/nepfil/serie-investigacao-filosofica/>

- origens do poder, da prosperidade e da pobreza. São Paulo: Elsevier.
- ALCHIAN, A. A. (1950). Uncertainty, evolution, and economic theory. *Journal of Political Economy*, 58(3), 211-221. <https://doi.org/10.1086/256940>
- APPLEBY, J. (1976). Locke, Liberalism and the Natural Law of Money. *Past & Present*, (71), 43-69. Retrieved July 18, 2020, from www.jstor.org/stable/650353
- ARIDA, P. (1996). A História do Pensamento Econômico como Teoria e Retórica. In Rego, J. M..*Retórica na Economia*. São Paulo: Editora 34.
- BOLTON, P., DESPRES, M., SILVA, L. A., Svartzman, R., & Samama, F. (2020). *The green swan: Central banking and financial stability in the age of climate change*.
- BROOME, J.&FOLEY, D. (2016). A world climate bank. In Axel Gosseries & Iñigo González-Ricoy (eds.), *Institutions for Future Generations*. Oxford University Press. pp. 156-169.
- CARTER, Z. D. (2020). *The price of peace: Money, democracy, and the life of John Maynard Keynes*. Nova Iorque: Random House.
- CHRISTIAN, B., & GRIFFITHS, T. (2017). *Algoritmos para viver: A ciência exata das decisões humanas*. Trad. Paulo Geiger. São Paulo: Editora Companhia das Letras.
- COASE, R. (1960). The Problem of Social Cost. *The Journal of Law & Economics*, 3, 1-44. Disponível em : <http://www.jstor.org/stable/724810>
- COASE, R. H. (1991). The Institutional Structure of Production. *Journal des Économistes etdes Études Humaines*, 2(4): 431-440. doi:10.1515/jeeh-1991-0401
- COWEN, T.(2018). *Stubborn attachments: A vision for a society of free, prosperous, and responsible individuals*.
- DELL, M.; JONES, B. F.; OLKEN, B. A. (2012). Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics*, 4 (3): 66-95.DOI: 10.1257/mac.4.3.66
- ERVIN, C.(2008). Sovereign Wealth Funds. *The OECD Observer*. **2008** (267): 21–22. doi:10.1787/observer-v2008-2-en.
- FRIEDMAN, M. (2014). *Capitalismo E Liberdade*. Trad. A. C. Serra. São Paulo: LTC.
- GARDNER, D.; TETLOCK, P. (2016). *Superprevisões: A arte E a ciência de antecipar o futuro*. Objetiva.

- GREENSPAN, A. (2013). *O mapa e o território: Risco, natureza humana e o futuro das previsões*. Translation: André Fontenelle e Otacílio Nunes. São Paulo: Penguin, Schwarcz.
- HARARI, Y. N. (2015). *Sapiens: Uma breve história Da humanidade*. L&PM.
- HARARI, Y. N. (2016). *Homo deus: Uma breve história do amanhã*. Editora Companhia das Letras.
- HUEMER, M. (2012). In Praise of Passivity. *Studia Humana*,1(2):12-28. Disponível em: <studiahumana.com/pliki/wydania/In%20Praise%20of%20Passivity.pdf>. Tradução: *Elogio da passividade*. Trad. Hélio S. C. Carneiro e Lucas Grecco. Disponível em: <<https://criticanarede.com/political.html>>
- JACOBS, A. J. (2018). Thanks a thousand: A gratitude journey. Nova Iorque: Simon & Schuster.
- JAWORSKI, P. (2020). *Bloody well pay them: the case for voluntary remunerated plasma collections*. London: Adam Smith Institute. Disponível em: <<https://adamsmith-private.squarespace.com/s/Bloody-Well-Pay-Them-Peter-Jaworski.pdf>>
- KAHNEMAN, D. (2012). *Rápido e devagar: duas formas de pensar*. Trad. Cássio de Arantes Leite. Rio de Janeiro: Objetiva.
- KEYNES, J. M. (1963). Economic Possibilities For Our Grandchildren. In: *Essays in Persuasion*. New York: W.W Norton & Co, 1963.
- KIRILENKO, A. A.; LO, A. W. (2013). Moore's Law vs. Murphy's Law: Algorithmic Trading and Its Discontents. *Journal of Economic Perspectives*, 27(2). doi:10.2139/ssrn.2235963
- LO, A. W. (2015). The Gordon Gekko Effect: The Role of Culture in the Financial Industry. Preprint in *SSRN Electronic Journal*. doi:10.2139/ssrn.2615625
- MULLAINATHAN, S., &SHAFIR, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.
- MYERSON, R. B. (2008). Perspectives on Mechanism Design in Economic Theory. *American Economic Review*, 98(3), 586-603. Nobel Lecture Prize do dia 8 de dezembro de 2007. doi:10.1257/aer.98.3.586
- NORDHAUS, W. (2007). A Review of the *Stern Review on the Economics of Climate Change*. *Journal of Economic Literature*, American Economic Association, vol. 45(3), pages 686-702, September. doi: 10.3386/w12741
- O'KEEFE, C., CIHON, P., GARFINKEL, B., FLYNN, C., LEUNG, J., & DAFOE, A.

- (2020, February). The Windfall Clause: Distributing the Benefits of AI for the Common Good. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 327-331).
- PARFIT, D. (2011). *On What Matters. Vol. .1* Oxford: Oxford University Press.
- PEDROSO, M. (2014). Racionalidade limitada e uso de informações técnicas em modelos de análise de políticas públicas: proposições sobre a perspectiva integradora da Análise Multicritério de Decisão Espacial Construtivista. *RP3 - Revista de Pesquisa em Políticas Públicas*. 10.18829/rp3.v0i2.10156.
- PERELMAN M. (2007) Taming Economists. In: *The Confiscation of American Prosperity*. Palgrave Macmillan, New York. https://doi.org/10.1057/9780230607064_12
- PERES, R. A. (2020). "Equality and Responsibility in Financial Crisis: an ethical approach to the regulation of bail-outs, moral hazards and accountability". *Working Paper Series* n. 520. Banco Central do Brasil. <<https://www.bcb.gov.br/pec/wps/ingl/wps520.pdf>>
- PIKETTY, T. (2014). *O capital no século XXI*. Trad. Arthur Goldhammer. São Paulo: Editora Intrínseca.
- PINKER, S. (2013). *Os anjos bons Da nossa natureza: Por que a violência diminuiu*. Trad. Laura Teixeira Motta. São Paulo: Companhia das Letras.
- PISTOR, K. (2017). Money's legal hierarchy. In Herzog, L. (ed.). *Just financial markets? Finance in a just society*. Oxford: Oxford University Press. doi:10.1093/oso/9780198755661.003.0008
- ROSLING, H.; ROSLING, O.; RÖNNLUND, A. R. (2019). *Factfulness: O hábito libertador de só ter opiniões baseadas em fatos*. Trad. Vitor Paolozzi. Editora Record.
- ROTH, A. E. (2007). Repugnance as a Constraint on Markets. *Journal of Economic Perspectives*, 21(3), 37-58. doi:10.1257/jep.21.3.37
- SANDBERG, J. (2016). Pension funds, future generations, and fiduciary duty. In Axel Gosseries & Iñigo González-Ricoy (eds.). *Institutions for Future Generations*. Oxford: Oxford University Press. pp. 385-399. doi: 10.1093/acprof:oso/9780198746959.003.0023
- SANDEL, M. (2012). *O que o dinheiro não compra: os limites morais do mercado*. Trad. Clóvis Marques. Rio de Janeiro: Civilização Brasileira.
- SEN, A. (2011). *A ideia de justiça*. Trad. Denise Bottman e Ricardo D. Mendes. São Paulo: Companhia das Letras.

- SORENSEN, R. A. (1988). *Blindspots*. Oxford England: Clarendon Press.
- SOROS, G. (2008). *Novo Paradigma para Os Mercados Financeiros*. Trad. Paulo Migliacci & Lúcia Boldrini. Rio de Janeiro: Ediouro.
- SPENCE, M. (2019). Marx against Marx: A critical reading of the Fragment on machines. *Communication, Capitalism & Critique*, 17(2), 327-339. <https://doi.org/10.31269/triplec.v17i2.1146>
- STERN, N. H. (2007). *The economics of climate change: The stern review*. Cambridge University Press.
- TALEB, N. N. (2015). *A lógica do Cisne Negro: O impacto do altamente improvável*. Trad. Marcelo Schild. Editora Best Seller.
- TIROLE, J. (2020). *Economia do bem comum*. Trad. André Telles. São Paulo: Zahar.
- WEATHERALL, J. O. (2015) *A física de Wall street*. Trad. Rio de Janeiro: Elsevier.
- WORLD HEALTH ORGANIZATION. (2015). *Progress on sanitation and drinking water : 2015 update and MDG assessment*. Disponível em: www.unicef.org/publications/index_82419.html >

Artigos da Stanford Encyclopedia of Philosophy

- HAUSMAN, D. M., "Philosophy of Economics", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/economics/>.
- FLEURBAEY, M., "Economics and Economic Justice", The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/economic-justice/>.
- DE BRUIN, B., HERZOG, L., O'Neill, Martin and Sandberg, Joakim, "Philosophy of Money and Finance", The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/money-finance/>.
- MORIARTY, J., "Business Ethics", The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/ethics-business/>

Filosofia da Religião

Luiz Helvécio Marques Segundo

1. Introdução

Imagine uma pessoa com extremos poderes, imensamente sábia, perfeitamente boa, que sempre existiu e nunca deixará de existir. Muito provavelmente você pensou em Deus, pois é assim que a tradição religiosa e cultural ocidental têm se referido a esse personagem. Deus é uma pessoa *onipotente, onisciente, sumamente boa* e que tem *existência necessária*. Considera-se, além disso, que um dos seus maiores atos de bondade tenha sido a criação do universo e dos seres humanos.

A crença em Deus é o elemento central das religiões abraâmicas, ou seja, do judaísmo, do cristianismo e do islamismo. Embora essas três religiões difiram em práticas e nas narrativas contidas em seus livros sagrados, elas têm o mesmo deus. A Torá, a Bíblia e o Alcorão reconhecem Deus como o criador do universo.

Em sentido lato, a expressão “crença religiosa” é usada para se referir a qualquer crença que envolva a aceitação de alguma afirmação que faça parte de alguma doutrina religiosa. Por exemplo, crença de Faruk de que Maomé é o profeta enviado de Deus, ou a crença de Saulo de que Jesus de Nazaré morreu para redimir o pecado da humanidade, são crenças religiosas. Neste texto, porém, “crença religiosa” será usada quase que exclusivamente para se referir à crença na existência de Deus. Nesse sentido, usaremos também a expressão “crença teísta”. Não que estejamos desprezando outras tradições religiosas, como por exemplo, as tradições afro-brasileiras; essa é apenas uma questão de restrição a uma discussão filosófica mais bem estabelecida. Não obstante, na medida em que as divindades de outras tradições se assemelharem ao deus abraâmico, problemas filosóficos similares também surgirão.

A crença religiosa -- i.e., a crença de que Deus existe -- é uma fonte de imensa reflexão filosófica. A mais intuitiva delas diz respeito às razões que poderíamos ter a favor dessa crença. Suponha que você acredite na existência de Deus, porque foi criado na igreja, ou talvez simplesmente porque assimilou a crença teísta em seu meio cultural. Nesse sentido, a sua crença religiosa é um tanto óbvia; você não coloca em xeque a existência de Deus. Suponha também que você tem um amigo que não acredita na existência de Deus; na verdade, ele nem faz ideia de quem pudesse ser Deus. Ele foi criado numa comunidade isolada, sem religião ou qualquer tipo de crença religiosa. Como você faria para convencê-lo racionalmente da existência de Deus? Pediria a ele que lesse a Bíblia? Ou tentaria por si próprio fornecer-lhe razões para acreditar que há um criador sumamente bom e extremamente poderoso?

Vamos supor que seu amigo seja inteligente e esteja prestes a iniciar uma agradável e cativante conversa sobre a existência de Deus. O restante deste texto servirá como roteiro para essa discussão -- faça um esforço em imaginar-se nessa conversa!

Na seção 2, é apresentado o conceito de Deus. Na seção 3, introduzimos três influentes argumentos a favor da existência de Deus: o argumento ontológico, o argumento cosmológico e o argumento do desígnio. Dois desafios à crença teísta são discutidos na seção 4: o problema do mal e o desacordo religioso. A seção 5, trata de um modelo epistemológico alternativo para a crença teísta. Por fim, na seção 6, veremos um argumento teísta contra o naturalismo filosófico, o principal oponente à visão de mundo teísta.

Pressuponho que o leitor tenha pouco ou nenhum contato com a discussão filosófica. Por isso apresento os problemas e argumentos ao longo do texto sem qualquer referência a textos ou autores. As referências e as indicações bibliográficas se encontram ao final do texto. O leitor verá a seguir versões simplificadas e informais de grandes ideias e argumentos influentes na história, antiga e recente, da filosofia.

2. A ideia de Deus

Uma boa discussão requer que os interlocutores estejam em sintonia. É preciso que estejam falando sobre a mesma coisa. É bastante decepcionante descobrir que, depois de duas horas de conversa, você e seu amigo estavam

falando de coisas completamente diferentes, embora parecesse o contrário. Por isso, vamos começar por deixar claro o que queremos dizer por “Deus”.

Vimos inicialmente que Deus é uma pessoa onipotente, onisciente, sumamente boa, que existe necessariamente. Essa pessoa, que é única, criou o universo. Mas o que realmente significa dizer que Deus tem essas propriedades?

Começemos pela ideia, ou conceito, de pessoa. Nós somos pessoas; e tendemos a considerar outros membros da nossa espécie como pessoas. Mas, é certo, não podemos dizer que “pessoa” e “membro da espécie *homo sapiens*” são o mesmo conceito, isto é, que representam a mesma ideia. A razão óbvia é que Deus, se existir, não pode pertencer à qualquer espécie biológica. Também diríamos que anjos, caso existam, são pessoas; e também considerariamos pessoas os membros de alguma raça extraterrestre que entrassem em contato conosco. Não é fácil, contudo, dizer no que consiste ser uma pessoa. Alguns se arriscariam a dizer que ser racional, ter consciência, usar uma linguagem sofisticada para se comunicar etc. são condições necessárias para que um indivíduo seja uma pessoa. Para os nossos propósitos, diremos que algo é uma pessoa, se pelo menos possui uma mente que lhe permite ter pensamentos, desejos e agir de acordo com eles.

Acontece que as pessoas com as quais estamos acostumadas são mentes incorporadas, isto é, são mentes “dentro” de um corpo. Mas Deus não possui um corpo; ele é uma mente desencarnada. A nossa mente possui uma íntima relação com o nosso cérebro, mas a mente de Deus não está conectada a qualquer cérebro. É preciso, portanto, conceber Deus como uma pessoa que é puramente mental (ou que é puro espírito) e que pode agir no e interagir com o mundo físico mesmo sem ter um corpo.

Passemos à onipotência. Dizer que Deus é onipotente é dizer que seus poderes de ação são ilimitados. Deus tem o poder de criar universos juntamente com as leis da natureza responsáveis por seu funcionamento. Ele pode também intervir no mundo quando e onde quiser. Embora eu tenha dito que os poderes de Deus são ilimitados, há na verdade alguns limites. O primeiro deles é lógico. Deus não pode fazer com que $2 + 2$ seja outra coisa que não 4; e também não pode fazer com que um triângulo tenha quatro lados; nem mesmo pode fazer com que um objeto seja ao mesmo tempo completamente verde e completamente vermelho. Deus também não pode se “suicidar”. Se ele é todo-poderoso, parece claro que ele pode usar esse poder para aniquilar a si próprio.

Mas também não dissemos que ele sempre existiu e sempre existirá? Se isso é verdade, então a onipotência divina esbarra aí.

Mas por que Deus tem de existir necessariamente? Não parece simples imaginar um mundo em que Deus não exista? Essa é uma noção bastante intrigante. Mas o que se pretende aqui é dizer que Deus não poderia não existir. Se o universo não poderia ter sido criado a não ser por Deus, como pensa o teísta, não faz sentido dizer que o universo existe, mas Deus não existe.

Dizer que Deus é onisciente significa que ele sabe tudo. Mas tudo o quê? Ele sabe todas as verdades sobre o mundo. Ao contrário dos seres humanos, que são limitados cognitivamente, Deus não é. Imagine um supercomputador com uma capacidade incrível de processamento e de memória; esse computador ainda seria inferior à mente de Deus. Lembre-se, não há qualquer empecilho físico (de hardware) à capacidade de processamento da mente de Deus (software).

Por fim, Deus é sumamente bom. A vontade de Deus é tal que ele não consegue agir a não ser para fazer o bem. Deus nunca agiria para fazer o mal. Como Deus é onisciente, ele sabe o que é moralmente bom e o que não é. Sendo assim, Deus nunca erra em seus juízos morais e, por conseguinte, nunca age a não ser de acordo com as regras morais corretas. Segue-se disso que Deus é moralmente perfeito.

Quadro 1: Quebra-cabeças envolvendo as propriedades divinas

O paradoxo da pedra

Se Deus é onipotente, então ele pode fazer (quase) tudo. Poderia então Deus criar uma pedra que fosse tão pesada que nem ele próprio pudesse levantar? Suponha que sim. Se ele criar tal pedra, então haverá alguma coisa que Deus não pode fazer, nomeadamente levantar essa pedra. Mas então suponha que não. Se ele não pode criar tal pedra, então há pelo menos uma coisa que Deus não pode fazer, nomeadamente, criar uma pedra tão pesada que nem ele próprio pudesse levantar. Seja como for, há pelo menos uma coisa que Deus não pode fazer; portanto, Deus não é onipotente. Você acha esse raciocínio convincente? Há alguma maneira de evitar a sua conclusão?

Presciência divina e livre-arbítrio

Se Deus é onisciente, então ele sabe todas as proposições verdadeiras sobre o mundo,

tanto proposições sobre o passado quanto proposições sobre o futuro. Alguém que sabe coisas sobre o futuro é presciente. A onisciência implica a presciência. Suponha que Deus saiba que você levantará a sua mão esquerda no dia 13 de abril de 2030, às 09:30 da manhã. Se ele sabe isso, então é verdade que você levantará a sua mão. Lembre-se que Deus não erra. Isso implica, então, que o seu futuro está determinado. Você não poderá fazer outra coisa no dia 13 de abril de 2030 às 09:30 a não ser levantar a sua mão esquerda. Por outro lado, fazer coisas como levantar a mão são ações que estão sob o nosso controle. Eu simplesmente escolho se vou levantar a mão ou não. Isso é o que significa dizer que tenho livre-arbítrio. Ora, se Deus sabe que no futuro você vai levantar a mão, então você não poderá fazer outra coisa que não levantar a mão. Você não terá escolha sobre isso. E se não terá escolha, não será livre. Generalizando, se Deus sabe o que vai acontecer no futuro da humanidade, isso implica que os humanos não poderiam agir de maneira diferente. Ou seja, implica que os seres humanos não têm livre-arbítrio. Portanto, se Deus é presciente, então não temos livre-arbítrio. Você concorda com essa conclusão? Você acha que se Deus existir nós não seremos pessoas livres?

3. Argumentos a favor da crença teísta

Depois de explicar ao seu amigo que tipo de entidade é Deus, ele certamente se sentirá curioso. Principalmente depois de perceber que mais de 90% de brasileiros -- mais de 189 milhões de pessoas -- possuem a crença na existência de Deus.

Ele se sentirá curioso principalmente porque Deus é um tipo de coisa bastante diferente das coisas com as quais ele está acostumado a lidar. Na comunidade em que ele cresceu as pessoas simplesmente tomavam a natureza como um dado, não viam a necessidade de explicar a sua existência. Além do mais eles tinham o seguinte costume por lá: se alguém fazia alguma afirmação sobre a existência de qualquer tipo de coisa extraordinária, tinha por obrigação oferecer as evidências de que dispunha para as outras pessoas. Eles acreditavam que se todos se envolvessem no exame dessas evidências, se pensassem cuidadosamente juntos, poderiam chegar a uma posição razoável.

Sendo você uma pessoa receptiva e uma boa anfitriã, certamente concordará em oferecer-lhe algumas evidências a favor da crença teísta. Talvez você resolvesse começar por oferecer-lhe alguns relatos da Bíblia, uma vez que você pensa que nesse livro sagrado estão contidas verdades importantes. Mas o

seu amigo poderia ter a seguinte reação: “Como sei que o que a Bíblia diz é verdade? Pelo que entendi, a escrita da Bíblia foi supostamente inspirada por Deus. Mas eu ainda não sei se é razoável acreditar em sua existência. Se não acredito em sua existência, como posso acreditar que ele inspirou o que quer que seja?”. Essa é uma reação sensata. O seu argumento não pode ser bom, uma vez que é um tipo de argumento que os filósofos chamam de *argumento circular*: ele pressupõe aquilo que deveria provar. Você terá de começar por outros argumentos. E aqui há uma tradição de argumentos com essa finalidade. Essa tradição é conhecida como “teologia natural”. É a tentativa de mostrar que Deus existe através da razão, sem o auxílio da inspiração ou revelação divina.

Quadro 2: Teologia Natural

A teologia natural faz parte de uma tradição intelectual bastante antiga. Podemos remontar os primeiros argumentos a favor da existência do deus teísta a Platão (428-348 AEC) e Aristóteles (384-322 AEC), na Grécia Antiga. O longo período medieval testemunhou uma produção riquíssima na teologia natural. Grandes pensadores cristãos como Anselmo de Cantuária (1033-1109), Tomás de Aquino (1225-74) e Duns Scotus (1266-1308); pensadores árabes como Ibn Sina (980-1037) e Ibn Rushd (1126-1198); e pensadores judeus como Maimônides (1138-1204) desenvolveram sofisticados argumentos a favor da existência de Deus, como também pensaram sobre a própria natureza de Deus.

No período moderno encontramos argumentos da teologia natural nas obras de René Descartes (1569-1650), John Locke (1632-1704), George Berkeley (1685-1753), Gottfried Leibniz (1646-1716) dentre muitos outros. Apesar da influência da teologia natural, vários filósofos foram críticos ao projeto de tentar provar que Deus existe. David Hume (1711-1776) e Immanuel Kant (1724-1804) foram dois dos grandes filósofos do período moderno a se queixar da teologia natural. Embora Hume fosse ateu, Kant era cristão.

No início do século XX, influenciados por Hume e Kant, os positivistas lógicos também faziam frente à teologia natural; eles defendiam que a existência de Deus – assim como várias outras questões tradicionais – não passavam de pseudo-problemas que surgiam pelo uso descuidado da linguagem.

Na segunda metade do século XX, porém, a filosofia da religião voltou a fazer parte das discussões “oficiais” novamente, sobretudo a teologia natural. Desde então, vários

filósofos têm-se dedicado a pensar sobre os velhos argumentos da teologia natural e a criar novos.

3.1. O argumento ontológico

O argumento ontológico é bastante peculiar; ele é uma tentativa de estabelecer a existência de Deus, usando apenas o pensamento, sem a necessidade de qualquer informação empírica. As suas principais premissas são *a priori*. Dizemos que uma proposição é *a priori* quando pode ser conhecida pelo pensamento apenas, sem a ajuda dos sentidos. Por exemplo, a proposição $2 + 2 = 4$ é conhecível apenas pelo pensamento; você não precisa usar os sentidos e coletar informação empírica para saber que é verdadeira. Os filósofos que usam esse argumento pensam que podemos saber *a priori* que Deus existe. Por isso diz-se que o argumento ontológico é um argumento *a priori*.

A estratégia aqui é que entender a ideia ou conceito de Deus nos leva a concluir que ele existe. Mas como? A maneira mais comum é através de uma forma argumentativa bem famosa: a *reductio ad absurdum* (que significa em português “redução ao absurdo”). Numa *reductio ad absurdum*, começamos com uma suposição e tentamos obter uma contradição, usando operações lógicas, feitas a partir da suposição e de um conjunto de premissas. Atingida a contradição, negamos a suposição. Isto é, dizemos que a suposição é falsa. A ideia é que se uma suposição leva a uma contradição, então é falsa.

Começaremos então pela suposição de que Deus não existe. Alguém que negue a existência de Deus, certamente entende o conceito de Deus. Não faz sentido alguém dizer “Papai Noel não existe” e não entender o conceito de Papai Noel; é preciso saber que é o velhinho morador da Lapônia que supostamente presenteia as crianças ricas no dia de Natal. Do mesmo modo, alguém que diz “Deus não existe” entende que está negando existência a um ser maximamente perfeito, um ser que é onipotente, onisciente, sumamente bom etc.

Vamos introduzir a seguinte distinção: coisas que existem no entendimento e coisas que existem na realidade. Você, as pirâmides do Egito, os gorilas na Montanha Virunga etc. são coisas que existem na realidade. Já o Papai Noel, Pégaso, o Pé Grande etc. são coisas que, embora não existam na realidade, existem no entendimento. Note que tudo o que existe na realidade,

uma vez que seja conteúdo do pensamento de alguém, tem também existência no entendimento. Mas nem tudo o que tem existência no entendimento tem também existência na realidade. O ateu, ao negar a existência de Deus, aceita que Deus exista no entendimento, mas não na realidade. Portanto, a suposição de que Deus não existe na realidade implica que Deus existe pelo menos no entendimento.

O conceito de Deus implica a sua máxima perfeição. Na verdade, estamos estipulando que ser *maximamente perfeito* é no mínimo ser onipotente, onisciente, sumamente bom etc. Assim, quando alguém entende o conceito de Deus, entende que Deus não pode ser menos do que maximamente perfeito. Podemos então perguntar: o que tem mais perfeição, um ser que existe apenas no entendimento ou um ser que também existe na realidade? Seria razoável dizer que um ser que existe na realidade é mais perfeito do que um que existe apenas no entendimento. Por exemplo, se o Papai Noel existisse na realidade, ele seria mais perfeito do que o Papai Noel existente no entendimento (ou nos contos de Natal!). Do mesmo modo, o ateu poderia dizer que se Deus existisse na realidade, seria mais perfeito do que se existisse apenas no entendimento.

Ora, muito cuidado aqui. Dissemos que Deus é por definição o ser maximamente perfeito. E o ateu acabou de dizer que o ser maximamente perfeito poderia ser mais perfeito. Mas se algo é maximamente perfeito, não poderia ser mais perfeito. Só algo menos que maximamente perfeito poderia ser mais perfeito. Assim, o ateu primeiro aceita que Deus é o ser maximamente perfeito por definição, e logo em seguida diz que Deus não é o ser maximamente perfeito. Eis a contradição.

Uma vez que a suposição de que Deus não existe na realidade nos leva a uma contradição, só nos resta negá-la. Mas negar essa suposição implica que a proposição de que Deus existe na realidade.

Os filósofos costumam organizar seus raciocínios de uma maneira peculiar; procuram separar as premissas das conclusões e deixar claro como cada passo do argumento conduz ao próximo. Esse procedimento facilita a avaliação do argumento. É importante que você comece a exercitar essa habilidade. Vamos esquematizar o argumento ontológico. Caso prefira, você pode pular essa parte e continuar a leitura no parágrafo seguinte (mas sugiro que volte mais tarde). Qualquer dúvida, consulte o capítulo sobre lógica. No esquema abaixo, a indicação se a proposição numa linha particular é uma

premissa ou uma conclusão será feita entre colchetes. Cada **Cn** representa uma conclusão intermediária; **C** representa a conclusão do argumento; os numerais separados por vírgula indicam quais foram as premissas usadas para obter aquele passo:

1. Deus não existe na realidade. [**Suposição**]
2. Deus é o ser maximamente perfeito. [**Definição**]
3. Tudo o que é conteúdo do pensamento tem existência no entendimento. [**Premissa**]
4. Deus existe no entendimento. [(**C1**): 2, 3]
5. Tudo o que existe na realidade é mais perfeito do que a sua contraparte no entendimento. [**Premissa**]
6. Se Deus existisse na realidade, seria mais perfeito. [(**C2**): 4, 5]
7. O ser maximamente perfeito poderia ser mais perfeito. [**6 paráfrase**]
8. Se algo poderia ser mais perfeito, então não é maximamente perfeito. [**Premissa**]
9. O ser maximamente perfeito não é maximamente perfeito. [(**C3**): 7, 8]
10. Logo, Deus existe na realidade. [(**C**): 1, *introdução da negação*]

Espero que seu amigo tenha admirado a bela peça de raciocínio que é esse argumento. Realmente engenhoso. Talvez ele o tenha achado difícil. Talvez ele até esteja se sentido ludibriado intelectualmente. Vale a pena percorrer cada passo do argumento e se convencer de que é dedutivamente válido. Em seguida, é importante investigar se há boas razões para se aceitar as premissas e as demais suposições do argumento.

Não convencido de que se pode provar a existência de algo tão substancial quanto Deus a não ser com auxílio de informações empíricas, ao mesmo tempo que está convencido da lógica impecável do argumento ontológico e de suas premissas, seu amigo resolve utilizar o seguinte recurso. Ele constrói um argumento semelhante ao argumento ontológico a favor da existência de Deus, porém substitui Deus por outra entidade, digamos a Ilha de Fidel. A Ilha de Fidel é, por definição, a ilha maximamente perfeita. Uma ilha sem miséria, com a

melhor a música, com as melhores pessoas etc. A Ilha de Fidel existe no pensamento, mas não na realidade. Ora, mas se por definição a Ilha de Fidel é maximamente perfeita, estamos a dizer que a ilha maximamente perfeita poderia ser mais perfeita... Logo, a Ilha de Fidel existe na realidade. (Fica como exercício ao leitor a reconstrução do argumento ontológico a favor da Ilha de Fidel).

Você deve ter notado que é possível construir um argumento ontológico para provar a existência de muitas coisas que só deveríamos acreditar depois de investigação empírica. Isso sugere que há algum pressuposto escondido no argumento que o torna implausível. Se você e seu amigo ainda tiverem fôlego, mãos à obra.

3.2. O argumento cosmológico

É natural que seu amigo, e mesmo você, tenha desconfiado do argumento ontológico. Há ainda outra categoria de argumentos da teologia natural, os argumentos *a posteriori*. Dizemos que uma proposição é conhecida *a posteriori* quando dependemos do uso dos sentidos, isto é, da experiência empírica, para justificá-la. Por exemplo, só podemos saber a verdade da proposição *Está chovendo lá fora*, recorrendo à experiência. É preciso olhar pela janela, ou sair à rua, ou perguntar a alguém, para justificar a crença de que está a chover lá fora. A reflexão sobre o argumento ontológico pode ter te convencido de que só se pode provar a existência de algo *a posteriori*.

É importante notar que argumentos *a posteriori* não excluem a possibilidade de premissas *a priori*. No entanto, a ocorrência de uma única premissa *a posteriori* num argumento é suficiente para que todo o argumento seja considerado *a posteriori*.

Dentre os mais antigos argumentos *a posteriori* a favor da existência de Deus está o argumento cosmológico. O argumento parte de uma observação empírica bastante geral: o universo existe. Essa premissa abrevia um conjunto imenso de coisas que existem: você, seu amigo, a sua cidade, as plantas, as rochas, os animais, os átomos que compõem cada coisa existente e assim por diante. Podemos considerar o universo como a totalidade de coisas que existem (e que existiram no passado, mas já não existem no presente).

A segunda premissa é mais abstrata. Ela diz que para todo o fato há

uma explicação para esse fato. Por exemplo, o que explica a existência de Brasília é um conjunto de outros fatos: vários prédios e ruas foram construídos no planalto central a mando do então presidente do Brasil, Juscelino Kubitschek, a fim de que se tornasse a capital do país. Esses outros fatos certamente também terão explicação. Os filósofos costumam dar um nome a essa exigência: Princípio de Razão Suficiente (PRS).

Podemos considerar três opções na explicação de um fato: (a) esse fato foi causado por outro fato, (b) esse fato é auto-existente ou (c) esse fato veio do nada. A opção (c) é ininteligível. Portanto, ou (a) um fato foi causado por outro ou (b) sempre existiu. Assim, ou um fato é dependente de outro para a sua explicação ou é autoexplicativo.

Nas premissas restantes, tenta-se estabelecer que as exigências explicativas impostas por PRS nos levam a postular a existência de Deus. Vamos abreviar a afirmação *O universo existe* com a letra “F”. De acordo com PRS, F precisa de explicação. Assim, ou (a) F foi causado por outra coisa ou (b) F é auto-existente.

Começemos por (b). Se o universo sempre existiu, então, em algum sentido, ele não poderia não ter existido. Mas facilmente podemos conceber a não existência do universo. Há algo, mas poderia nada haver. A existência do universo é um fato contingente. Fatos contingentes não são auto-explicativos. Portanto, F tem de ser explicado pela existência de outra coisa.

Passemos a (a). Que tipo de fato explica F? Aqui as coisas começam a ficar ainda mais abstratas. Suponhamos que há um outro fato, F_1 , que explica a existência de F. De acordo com PRS, ou (a) F_1 é explicado por outros fatos, ou (b) é auto-explicativo. Note, porém, que qualquer que seja esse F_1 , ele fará parte do conjunto da totalidade das coisas que existem; por conseguinte, fará parte do universo. Se, portanto, F_1 depender de outros fatos, teremos de aumentar ainda mais o nosso conjunto de coisas que existem e novamente poderemos perguntar porque essas coisas existem ao invés de não existir. Esse processo pode ser repetido infinitamente, de modo que nunca responderíamos genuinamente à pergunta “Por que há algo, e não nada?”. Esse regresso infinito de perguntas só pode ser interrompido quando encontrarmos um fato auto-explicativo que explique a existência do universo.

Talvez você tenha notado que não dissemos qual é o tamanho da cadeia explicativa que leva do F_n auto-explicativo a F. Poderia ser uma cadeia

enorme, como também poderia ser brevíssima, terminando em F_1 . Há alguma razão para determinarmos o tamanho dessa cadeia? Do ponto de vista teísta, sim. Postular a existência de Deus como F_1 é a melhor explicação para F . Deus sempre existiu e, por algum motivo (provavelmente ligado à sua suma bondade), resolveu criar todo o restante do universo.

Assim como fizemos na seção anterior, com o argumento ontológico, esquematizemos o argumento cosmológico como se segue:

1. O universo existe. **[Premissa]**
2. PRS. **[Premissa]**
3. Há uma explicação para a existência do universo. **[(C1): 1, 2]**
4. Ou a existência do universo é auto-explicativa ou há um F_1 que a explique. **[Premissa/PRS]**
5. A existência do universo não é auto-explicativa. **[Premissa]**
6. Há um F_1 que explica a existência do universo. **[(C2): 4, 5]**
7. F_1 é auto-explicativo ou há um F_n que explique F_1 . **[6, PRS]**
8. Se F_1 não for auto-explicativo, então ou uma cadeia explicativa infinita é gerada ou uma cadeia maior é gerada. **[PRS]**
9. Não pode haver uma cadeia infinita de explicações. **[Premissa]**
10. Uma cadeia menor é preferível a uma maior. **[Premissa]**
11. F_1 é auto-explicativa. **[(C3): 7, 8, 9, 10]**
12. Logo, F_1 é a existência de Deus. **[(C): Melhor explicação]**

É importante esclarecer que a existência de Deus não foi introduzida para explicar cada fato particular. Por exemplo, o fato de a porção de água que João esquentou ter entrado em ebulição a 71°C é explicado, dentre outras coisas, pelo fato de João estar no alto do Monte Everest, cuja altitude é elevadíssima, e pelas leis da termodinâmica, tem uma influência na temperatura da ebulição da água. É claro que o fato de João estar no Monte Everest pode ser explicado por outros fatos e esses fatos, por sua vez, podem ser explicados por outros e assim por diante. Ainda que uma cadeia causal retroceda até a existência de Deus, daí não se segue que a existência de Deus seja a explicação para a porção de água de João ter entrado em ebulição a 71°C . Pensar assim

seria um equívoco. Os filósofos costumam dizer que a explicação não é transitiva: mesmo que A explique B e B explique C, daí não se segue que A explique C. Você provavelmente conhece algumas relações não transitivas: João ama Teresa e Teresa ama Raimundo mas daí não se segue que João ame Raimundo. Considere outro caso de explicação. Suponha que a planta que tenho em minha escrivaninha morreu. A explicação é que ela não estava recebendo luz. E isso é explicado pelo fato de eu ter me ausentado de casa por meses. O que por sua vez é explicado pelo fato de eu estar viajando a trabalho. E o que explica minha viagem é o fato do meu patrão confiar em mim. Quer dizer então que o fato de o meu patrão confiar em mim explica por que a minha planta morreu? Obviamente que não. Do mesmo modo, introduzir a existência de Deus no final de uma longa cadeia explicativa não implica que a existência de Deus explique o primeiro fato, embora explique porque há toda a coleção de fatos. A existência de Deus é introduzida para explicar o fato geral de o universo existir. Ou seja, a existência de Deus explica o fato geral de existirem fatos.

É comum encontrar críticas ao argumento cosmológico dizendo que é um tipo de falácia. Diz-se que o teísta parte da premissa de que todas as coisas possuem uma causa (ou explicação) e conclui que há uma causa que é a causa de todas as coisas. Isso seria como argumentar que só porque todas as pessoas possuem uma mãe, há uma mãe que é a mãe de todas as pessoas. Esse tipo de raciocínio é de fato falacioso. Seria prudente que você alerte seu amigo que o argumento cosmológico não tem essa forma. A conclusão de que Deus é a explicação da existência do universo decorre das exigências de explicação do PRS e da ideia de que Deus é a melhor explicação.

Por outro lado, seu amigo poderia dizer que postular a existência de Deus é desnecessário, uma vez que explicar cada um dos fatos que compõem o universo significa o mesmo que explicar a existência do próprio universo. Por exemplo, se você explica a existência de cada uma das plantas e dos animais da floresta amazônica, você também está explicando a existência da floresta; não parece ser preciso postular um fato adicional que explique a existência da floresta.

Mas e se quisermos saber por que há uma floresta ali ao invés de outra coisa? Por exemplo, por que não há ali um deserto? Esse fato certamente não é explicado pela explicação da existência de cada um dos elementos que compõem a floresta. A explicação depende de fatores ambientais, climáticos etc.

que existiram bem antes da própria floresta. O mesmo vale para o universo. Explicar cada um dos fatos que compõem o universo, portanto, não é a mesma coisa que explicar a existência do universo.

O seu amigo, contudo, poderia perguntar pela plausibilidade do PRS. Afinal de contas, por que tem de haver uma explicação para todo fato? Não poderia haver fatos brutos, isto é, fatos sem qualquer explicação? E não poderia ser a existência do universo um desses fatos brutos?

Ou ainda, mesmo que PRS seja verdadeiro, o que garante que tenhamos de compreender ou mesmo conceber todas as explicações possíveis? Não seria muita pretensão pensarmos que somos capazes de compreender todo o universo? Afinal, há milhares de anos a espécie humana vem tentando compreendê-lo; e, nesse empreendimento, o erro tem sido a tônica! Talvez o teísta tenha deixado escapar uma distinção importante: entre uma componente metafísica e uma epistêmica do PRS. A componente metafísica diz respeito aos fatos e a componente epistemológica diz respeito à nossa capacidade de conhecer os fatos explicativos. Fazendo essa distinção, podemos ver que o PRS metafísico não implica o PRS epistêmico. Considere o seguinte caso. Um cego de nascença procura entender por que algumas pessoas gostam de ir a exposições de pinturas abstratas. A explicação é que elas gostam de observar as cores e formas das pinturas. O cego não possui o conceito de cores, de modo que ele não é capaz de entender a explicação; muito embora ele saiba que há tal explicação. O mesmo poderia se passar conosco ao tentar entender o universo. Ainda que seja verdade que haja um fato que explique a existência do universo, daí não segue que sejamos capazes de entendê-los. E talvez o nosso caso seja ainda pior, uma vez que, ao contrário do cego, que ainda tem uma indicação da existência de pinturas, talvez nem tenhamos a indicação daquilo que explique o universo.

Tenha seu amigo se convencido ou não, o argumento cosmológico é uma bela peça de raciocínio e merece ser avaliado com cuidado.

3.3. O argumento do desígnio

Passemos ao último argumento do nosso pequeno tour pela teologia natural: o argumento do desígnio. Diferentemente do argumento cosmológico, o argumento do desígnio parte de uma observação mais bem delimitada: a

existência de ordem e complexidade no universo. Essa ordem está geralmente ligada a algo que poderíamos chamar de adequação de meios a fins encontrada no universo. Por isso é também comum chamar esse tipo de argumento de “argumento teleológico” (“telos” quer dizer em grego “fim”, “finalidade”).

O argumento do desígnio foi bastante influente nos séculos XVIII e XIX, sobretudo devido à situação em que a biologia (chamada de “história natural” à época) se encontrava na época. Hoje em dia há um movimento denominado “desígnio inteligente”, que dentre outras coisas defende o ensino do criacionismo nas aulas de biologia. Esse movimento, embora faça uso da teologia natural, é mais um movimento político do que intelectual. O nosso interesse aqui, vale notar, é intelectual. Além do mais, ao contrário do movimento político do desígnio inteligente, a teologia natural não está em conflito com a ciência. Passemos ao argumento.

Suponha que você esteja passando por um ferro velho e veja um carro em perfeitas condições de funcionamento. Além disso, esse carro é de um modelo exótico, não disponível no mercado. Certamente você se perguntará o porquê de aquele carro estar ali. Afinal, tudo o que esperamos encontrar num ferro velho são peças velhas, latarias amassadas etc.

Para que um carro funcione é preciso que cada peça esteja no lugar certo, desde o motor até a barra de direção. Cada uma dessas peças e mecanismos foram projetados para uma finalidade principal: a locomoção do carro. Se você encontra um carro em condições de funcionamento num lugar onde você não esperava encontrar um, então ou alguém o deixou lá ou um mecânico o montou a partir de boas peças encontradas ali no ferro velho. Dado o padrão das peças e da lataria, é razoável inferir que alguém inteligente construiu aquele carro ali a partir de peças retiradas de outros carros. O que não é razoável inferir é que esse carro tenha aparecido ali por um processo completamente aleatório de combinação de peças -- *e.g.*, um tufão ter passado e levantado todas aquelas peças que, depois de se debaterem umas contra as outras, acabaram se encaixando e formando um carro em perfeitas condições de funcionamento. Não! A conclusão óbvia aqui é que há um mecânico.

Considere agora a diversidade de seres vivos existentes no globo terrestre. Cada um deles está muito bem adaptado a seus meios; parecem ter sido feitos para viver ali. Possuem mecanismos tão complexos (e alguns ainda mais) quanto ao de um motor de um carro. O olho humano, por exemplo. É um

dispositivo cuja função é captar informação do ambiente externo do organismo. Essa informação é enviada primeiramente em forma de luz; o olho é como uma câmera. Só depois é que ela é convertida em sinais eletroquímicos e chega até o cérebro, onde é traduzida em percepção. Mas, para que a captação da luz seja por fim traduzida em percepção, é preciso que o olho esteja funcionando adequadamente. Cada uma de suas partes tem de estar ajustada. Por exemplo, o sistema de lentes (córnea/humor-aquoso/cristalino) tem de ser tal que capte o feixe de raios luminosos divergentes e o convirja para um ponto no fundo do olho (a retina); a íris tem de controlar a intensidade de luz que entra no olhos (dilatando ou contraindo a pupila); os músculos ciliares têm de controlar a forma do cristalino a fim de determinar a vergência do olho; por fim, a informação óptica, na retina, tem de ser traduzida em um impulso nervoso que é transmitido, via nervo óptico, até o encéfalo. E assim como o motor de um carro, qualquer coisa mal regulada ou fora do lugar compromete o funcionamento do mecanismo.

Quando consideramos, na natureza, mecanismos como o olho, a curiosidade que tivemos no ferro velho ressurgiu: que tipo de processo foi responsável pela sua produção? Seria razoável esperar que a natureza os tivesse produzido através de processos aleatórios? Que, assim como no ferro velho, as peças mais simples estivessem espalhadas e alguma força natural as tenha juntado ao acaso de modo que formassem um mecanismo complexo com funções específicas? Se o raciocínio anterior, sobre o surgimento do carro no ferro velho, nos levou a concluir que havia um mecânico ou um engenheiro inteligente responsável pelo carro, também o mesmo tipo de raciocínio nos deveria levar a concluir que há um tipo de mecânico ou engenheiro responsável por mecanismos biológicos como o olho. Esse projetista, pensa o teísta, é Deus.

Resumindo: Há complexidade no mundo vivo (biológico). Qual é a melhor explicação (a mais plausível) para esse fato? A aleatoriedade ou a intenção de um projetista poderoso e sábio? É implausível pensar que algo como o olho humano tenha sido produto do acaso, assim como um automóvel não pode ter sido produto do acaso. A melhor explicação é a intenção de um projetista poderoso e sábio (Deus). Portanto, provavelmente Deus existe.

Esse argumento parece poderoso. Sem os desígnios de Deus não seríamos capazes de explicar as complexidades do mundo vivo. Mas será que só há essas duas hipóteses explicativas? Certamente que não. A biologia, desde a segunda metade do século XIX, dispõe de uma explicação plausível e

consensual: a evolução biológica. Mais especificamente, a teoria da seleção natural. O processo de seleção natural – que não é um processo aleatório – atua sobre os organismos ao longo do tempo e é responsável por grande parte da complexidade das funções biológicas encontradas na natureza.

Mas como a evolução biológica produz algo como o olho humano? Para entendermos isso, primeiro é preciso falar um pouco mais da evolução. Primeiro, “evolução”, na biologia, que dizer nada mais do que mudança. Dizer que uma espécie evolui quer dizer apenas que a frequência de certos genes (que determinam certas características fenotípicas) foi alterada ao longo de gerações. Segundo, há mecanismos responsáveis pela evolução, sendo o principal deles a seleção natural. Três coisas precisam ocorrer para que haja seleção natural: (i) variação de caracteres entre os indivíduos da população, (ii) a posse de certos caracteres favorecer a sobrevivência dos indivíduos naquele meio e (iii) hereditariedade. Um exemplo nos ajuda a compreender. Vamos considerar o famoso caso da mariposa sarapintada (*Biston Betularia*). Antes da revolução industrial no século XIX, em Manchester, Inglaterra, era comum a ocorrência de duas variedades de mariposas sarapintadas, uma de coloração mais clara e a outra mais escura. A coloração é determinada por um conjunto de genes. Por causa do aumento da poluição, causada pela atividade industrial, as árvores em Manchester começaram a ficar com os troncos mais escuros. As mariposas mais escuras podiam então se camuflar com facilidade nessas árvores, o que diminuía as chances de serem predadas pelos pássaros. As mais claras, por outro lado, eram predadas com mais facilidade. Assim, nesse meio, ter cor escura dava uma vantagem de sobrevivência; as mariposas com essa coloração conseguiam, portanto, chegar à idade adulta e se reproduzir, passando adiante os genes para a coloração escura. As mariposas claras morriam mais e, por conseguinte, reproduziam-se menos. Como consequência, a frequência de genes para coloração escura passou a ser predominante naquelas populações de mariposas. Assim, ao longo de algumas gerações, através do processo de seleção natural, a natureza operou uma mudança na espécie.

O leitor atento poderia dizer: Deus criou mariposas claras e escuras; apenas calhou de o ambiente não favorecer as claras. A seleção natural pode operar sobre a diversidade criada por Deus. Além do mais, ainda não vislumbramos como a evolução produziu o olho. Do mesmo modo, Deus pode ter criado criaturas com olhos e calhado delas terem se dado bem no ambiente.

Pois bem, aqui há outro aspecto da teoria da evolução que é preciso considerar, a *árvore da vida*. Todas as espécies atualmente existentes e todas aquelas que existiram no passado estão relacionadas por ancestralidade. Por exemplo, todas as espécies humanas (do gênero *homo* e da qual só o *homo sapiens* ainda não foi extinta) e os chimpanzés possuem um ancestral comum que viveu há cerca de 7,5 milhões de anos. Todos os primatas (incluindo nós, os chimpanzés, os lêmures, etc.) possuem um ancestral comum, um mamífero, que viveu há cerca de 65 milhões de anos. Os mamíferos, por sua vez, possuem um ancestral comum com répteis e aves, que viveu há 340 milhões de anos. Se retrocedermos mais uns 3 bilhões de anos, encontraremos um ancestral comum de todas as espécies que habitam e já habitaram a face da terra.

Toda característica fenotípica de uma espécie possui uma história evolutiva. Em geral, é a história da modificação de determinada estrutura que foi sendo feita ao longo de gerações e mais gerações. Considere, por exemplo, as mãos dos humanos, as nadadeiras da baleia e as asas do morcego. Aparentemente elas nada têm a ver, mas se seguirmos suas histórias evolutivas, veremos que são modificações de uma mesma estrutura de membros anteriores do ancestral comum dos tetrápodes -- essencialmente os mesmos ossos, porém com tamanhos, formas e ligamentos diferentes. O acúmulo de pequenas modificações em certas estruturas ao longo de gerações gera mudanças substanciais; a evolução, em geral, é gradual. Além do mais, são esses mesmos acúmulos de pequenas modificações, juntamente com outros fatores -- *e.g.*, isolamento geográfico e reprodutivo -- os responsáveis pela *especiação*, *i.e.*, o surgimento de novas espécies.

Por fim, ao olho. O olho humano é uma "invenção evolutiva" gradual, é o resultado do acúmulo de modificações em uma estrutura ancestral. Se você está pensando no surgimento do olho como um processo evolutivo de um único passo, esqueça! Nenhum biólogo pensa que estruturas complexas nos seres vivos aparecem do dia para a noite. Os biólogos acreditam que o olho tenha se desenvolvido a partir de uma estrutura capaz de captar luz, não servindo como um órgão visual, em proto-vertebrados que viveram há 550 milhões de anos.

Levando-se em conta, portanto, a teoria da evolução, podemos ver que o argumento do desígnio depende de uma premissa falsa: a de que a complexidade do mundo natural é explicada ou pela aleatoriedade ou pela intenção (desígnio) de Deus.

Cabe notar, ademais, que a teoria da evolução não é apenas uma hipótese explicativa, dentre muitas outras igualmente plausíveis, para a diversidade e complexidade do mundo vivo. A teoria da evolução é uma das pedras de toque de toda a biologia moderna. Não é estratégia sábia, portanto, recusar uma teoria científica bem estabelecida só para mantermos uma conclusão filosófica que nos agrada.

Há, é claro, maneiras de se reformular o argumento do desígnio. Uma delas seria procurar por fenômenos biológicos cuja complexidade seja irreduzível à explicação evolutiva. Alguns filósofos e cientistas teístas pensam ter encontrado essa complexidade irreduzível em fenômenos bioquímicos. Outra maneira é procurar outros fenômenos naturais que supostamente indicam desígnio e intencionalidade. Alguns pensam tê-los encontrado na própria organização do universo; argumentam que o ajuste fino existente no universo, que é expresso nas constantes cosmológicas, só pode ser explicado pela inteligência divina.

Quadro 3: A teoria da evolução

A teoria da evolução, tal como é hoje aceita na biologia, remonta ao trabalho de dois naturalistas britânicos, Charles Darwin (1809-1882) e Alfred Russel Wallace (1823-1913). Ambos chegaram independentemente à ideia de seleção natural, o principal mecanismo responsável pela evolução biológica. Em 1858 Wallace enviara um manuscrito a Darwin, “Sobre a tendência das variedades de se afastarem indefinidamente do tipo original”, que antecipava as suas ideias sobre a evolução e a seleção natural. Ainda em 1958 eles fizeram uma apresentação conjunta na Sociedade Lineliana de Londres. No ano seguinte Darwin publica *A Origem das Espécies*, umas das obras mais importantes da história intelectual humana.

Darwin e Wallace não foram os primeiros evolucionistas. Por exemplo, os naturalistas franceses, Buffon (1707-1788) e Lamarck (1744-1829), já tinham defendido que o conjunto de seres vivos na Terra sofriam modificações ao longo do tempo. Buffon acreditava que a natureza criava diferentes formas, de diferentes complexidades, e que, sob influência do ambiente, essas formas poderiam variar e dar origem a novas formas. Era como se as espécies tivessem um molde interno que permitisse alguma variação. Mas essa variação, pensava Buffon, era limitada. Lamarck, por sua vez, pensava que a diversidade da vida se dava pela modificação de tipos menos complexos (primitivos) até tipos mais complexos. Ele acreditava que a evolução fosse um processo linear que

trabalhasse em prol da complexidade dos organismos, que surgiam constantemente por geração espontânea. O ambiente desempenhava um papel secundário no processo: diferentes ambientes obrigavam os organismos a usar certas partes do corpo, o que poderia produzir alguma alteração; partes não usadas tendiam a se atrofiar. Essas mudanças eram transmitidas para a geração posterior. (Daí o mote pelo qual Lamarck é geralmente lembrado, as teorias do uso e desuso e da transmissão de caracteres adquiridos. Essas ideias que eram comumente aceitas na época, não sendo nenhuma novidade e nem de importância central na teoria de Lamarck).

O que separa Darwin e Wallace dos evolucionistas precedentes são as ideias de seleção natural e da árvore da vida. A seleção natural implica que a evolução não é um processo linear, mas antes um processo de divergência a partir de ancestrais. Essa divergência produz toda a variedade de espécies de que temos notícias. Cada nova espécie é uma ramificação da árvore da vida. É importante ressaltar que a teoria da evolução não é uma teoria sobre a origem da vida. Darwin e Wallace não procuraram explicar como surgiu o primeiro ser vivo. Eles simplesmente tomam como ponto de partida a existência da vida; havendo vida, basta que haja variedade numa população e pressão do meio para que ocorra seleção natural. A seleção natural explica a variedade da vida!

A teoria moderna da evolução, ou síntese moderna, foi estabelecida nas décadas de 30 e 40 do século XX por biólogos como Ronald Fisher (1890-1962), Theodosius Dobzhansky (1900-1975), Ernst Mayr (1904-2005), dentre outros. A síntese consistiu na combinação da teoria darwinista da seleção natural com a genética mendeliana. Isso não resultou apenas num modelo matemático para tratar a transmissão de caracteres, mas também permitiu a unificação de várias áreas da biologia em torno da evolução; nas palavras do próprio Dobzhansky, “na biologia nada faz sentido exceto à luz da evolução”.

4. Argumentos contrários à crença teísta

Depois de um intenso esforço intelectual na tentativa de compreensão e avaliação dos argumentos teístas, seu amigo lhe procura, uma semana depois, dizendo ter descoberto algumas razões para não acreditar em Deus. Uma dessas razões, que ele considera bastante forte, é a existência do mal no mundo. A outra, não tão forte, é a existência da pluralidade de religiões e os desacordos entre os próprios religiosos.

O seu amigo pensa que essas razões, ainda que não sejam fortes o bastante para apoiar a crença de que Deus não existe, são forte o suficiente para

que suspendamos a crença religiosa. A posição que recomenda a suspensão da crença religiosa chama-se “agnosticismo”.

4.1. O mal e a existência de Deus

Uma rápida espiada ao mundo à nossa volta releva uma quantidade enorme de males. O principal o tipo de mal é o sofrimento, tanto humano quanto não-humano. No momento em que escrevo, o mundo passa por uma pandemia de covid-19 -- que pode causar uma síndrome respiratória aguda e levar o indivíduo à morte -- causada por um vírus (o Sars-Cov-2). Até agora, mais de 6 milhões em todo o mundo foram contaminadas e mais de 400 mil pessoas morreram. Essa é só mais uma de várias pandemias registradas na história da humanidade. Some-se a isso, guerras – duas grandes guerras mundiais e centenas de guerras regionais e civis. Ditaduras, torturadores, psicopatas, violência gratuita, escravidão, fome, miséria, falta de saneamento, devastação do meio-ambiente e assim por diante. Depois de uma lista dessas é difícil não se desesperar. Sobretudo, porque a nossa intuição diz que um ser sumamente bom não desejaria que suas criaturas sofressem, e podendo eliminar esse sofrimento, ele o faria. Mas então por que Deus não elimina todo esse mal?

4.1.1. O argumento lógico

A intuição acima pode ser expressa dizendo-se que a existência de Deus é logicamente incompatível com a existência de mal no mundo. Isso quer dizer que as duas coisas não podem co-existir; ou Deus ou o mal, mas não ambos. Desse ponto de vista, a existência de Deus pode ser refutada através de uma consideração lógica. Quando duas proposições são logicamente incompatíveis, ambas não podem ser verdadeiras ao mesmo tempo. Assim, uma vez que é verdadeiro que o mal existe, tem de ser falso que Deus existe.

Vejamos passo a passo. Considere o seguinte conjunto, C, de proposições:

- I. Deus é existe.
- II. Há males no mundo.

Se a existência de Deus e do mal são incompatíveis, então o conjunto C tem de ser inconsistente, isto é, tem de ser possível deduzir uma contradição a partir das proposições I e II.

Contradições são representadas como " $p \& \text{não-}p$ ". E se olharmos para C, não há tal proposição lá. Além disso, a contradição não pode ser deduzida a partir apenas de I e II. Nesse caso, é preciso considerar se há alguma proposição implícita em C e que em conjunção com I ou II produz uma contradição. É razoável pensar que a seguinte proposição está implícita, uma vez que decorre do conceito de Deus:

III. Se Deus existe, então é onisciente, onipotente e sumamente bom.

Mas essa proposição ainda não é suficiente para obtermos a contradição. Ela no máximo nos permite deduzir de I e III:

IV. Deus é onisciente, onipotente e sumamente bom.

É preciso introduzir mais uma. Vejamos:

V. Se Deus é onisciente, onipotente e sumamente bom, então ele elimina todo o mal do mundo.

Agora, de IV e V segue-se por *modus ponens* que Deus elimina todo o mal do mundo e, por conseguinte, que:

VI. Não há males no mundo.

E agora, conjuntando II e VI, temos a contradição:

VII. Há e não há males no mundo. ($p \& \text{não-}p$)

Se um conjunto de proposições é inconsistente, então pelo menos uma de suas proposições originais é falsa. Uma vez que II é verdadeira, I tem de ser falsa. Lembre-se que as proposições III e IV estão sendo consideradas implícitas em C. As proposições originais são apenas I e II.

Por uma questão de simples exame lógico do conjunto de crenças do teísta, o seu amigo foi capaz de mostrar que Deus não existe.

Antes de ceder a seu amigo a pena pensar sobre as proposições supostamente implícitas em C. Os lógicos costumam dizer que uma proposição está implícita num conjunto apenas quando essa proposição puder ser deduzida daquelas proposições que são explícitas. A proposição III pode ser explicitada a partir I e da definição de Deus. É claro que você poderia negar III, mas ao custo de sustentar que ou (a) Deus não é onipotente ou (b) Deus não é onisciente ou (c) Deus não é sumamente bom. Mas será que um teísta estaria disposto a aceitar um Deus menos que maximamente perfeito?

Passemos a V. Ela não pode ser explicitada no sentido lógico, uma vez que não pode ser deduzida das proposições já explícitas em C. Mas há uma outra maneira de considerarmos se uma proposição está ou não implícita num conjunto. É quando essa proposição expressa uma verdade necessária. A melhor maneira de enxergar isso é através de um exemplo. Considere o conjunto, D, de proposições:

- a. João é solteiro.
- b. João é casado.

D é claramente um conjunto inconsistente. a e b não podem ser verdadeiras ao mesmo tempo. Porém, assim como C, D não contém uma contradição explícita. Temos de introduzir:

- c. Se x é solteiro, então x é não-casado.

E da conjunção de a e c, obtemos

- d. João é não-casado.

E da conjunção de b e d temos a contradição

- e. João é casado e não-casado.

Note que a proposição c não poderia ser falsa, ela é necessariamente

verdadeira. Você pode pensar em várias maneiras diferentes como o mundo poderia ser; em nenhuma delas será falso que todo solteiro é não-casado. Há circunstâncias, ou mundos possíveis, em que João não é solteiro; nesses mundos ainda assim não vai ocorrer de alguém ser solteiro e não-casado ao mesmo tempo. Sendo assim, faz sentido dizer que a proposição *c* está implícita no conjunto *D*.

Mas será que a proposição *V* é necessariamente verdadeira? Para avaliarmos essa proposição é preciso introduzir um pouco de terminologia nova, a terminologia dos mundos possíveis. Essa terminologia nos ajuda a raciocinar com um pouco mais de fluência sobre proposições precedidas pelos advérbios “necessariamente” e “possivelmente” ou pelas locuções “é necessário” e “é possível”. Um mundo possível é um modo como as coisas podem ser; é como se fosse uma história alternativa da realidade atual. O modo como as coisas são, a história da nossa realidade, é chamado de “mundo efetivo”¹⁴⁴. O mundo efetivo é também um mundo possível. Diremos que uma proposição é *necessariamente verdadeira* quando for *verdadeira em todos os mundos possíveis* (incluindo o mundo efetivo). Por exemplo, a proposição *Todo solteiro é não-casado* é verdadeira em todos os mundos possíveis. As proposições que são verdadeiras no mundo efetivo, mas falsas em pelo menos um mundo possível são proposições possivelmente falsas; por exemplo, a proposição *Michael Jackson é solteiro*. E as proposições que são falsas no mundo efetivo, mas verdadeiras em pelo menos um mundo possível são proposições, embora efetivamente falsas, possivelmente verdadeiras; por exemplo, a proposição *Michael Jackson está vivo*.

Será que Deus elimina todo o mal em todos os mundos possíveis? O teísta obviamente pensa que não, afinal o mundo efetivo contém uma quantidade substancial de males. E o teísta tem uma boa razão para pensar que a proposição *V* não é necessariamente verdadeira; vamos chamá-la de “defesa do livre-arbítrio”.

Segundo o defensor do argumento do mal: *necessariamente*, se Deus é onisciente, onipotente e sumamente bom, então ele elimina todo o mal do mundo. Ou seja, a proposição *V* é verdadeira em *todos os mundos possíveis*. Se houver, contudo, *pelo menos um* mundo possível em que Deus é onisciente, onipotente e sumamente bom, e ainda assim não elimine o mal, a proposição *V* não será

¹⁴⁴ Ou “mundo atual”, tal como o chamam alguns outros filósofos.

necessariamente verdadeira. O teísta só precisa mostrar que é *possível* que Deus não elimine todo o mal; isto é, que há pelo menos um mundo possível em que Deus existe e permite o mal.

Quadro 4: Necessidade e mundos possíveis

Já falamos aqui sobre a existência *necessária* de Deus e agora falamos sobre uma proposição *necessariamente* verdadeira. A maneira pela qual traduzimos as proposições prefixadas pelos advérbios modais “necessariamente” e “possivelmente” (ou pelas locuções “é necessário que” e “é possível que”) foi através do jargão dos mundos possíveis.

“Necessariamente p ” significa “Em todos os mundos possíveis p é verdadeira”, e

“Possivelmente p ” significa “Em pelo menos um mundo possível p é verdadeira”.

Um mundo possível é uma descrição completa da realidade, isto é, é um conjunto de proposições que descrevem a realidade. Os lógicos usam o termo “conjunto maximal consistente” de proposições. Isso quer dizer que, nesse conjunto, para qualquer proposição A , ou A pertence a esse conjunto ou $\neg A$ pertence; e é consistente porque não possui contradições. Qualquer conjunto maximal consistente de proposições é um mundo possível. O conjunto de proposições que descreve o mundo efetivo é também um mundo possível (afinal, não poderia ser impossível!).

Muitos lógicos acreditam que proposições são entidades abstratas. Isso quer dizer, então, que um mundo possível é uma entidade abstrata? Se não, será então uma entidade concreta? Ou ainda, faz sentido dizer que existem mundos possíveis? Essa não é uma discussão que precisamos resolver para usar o jargão dos mundos possíveis. A discussão sobre a natureza dos mundos possíveis é uma discussão metafísica. Usar o jargão dos mundos possíveis é uma questão heurística -- *i.e.*, facilita a nossa maneira de raciocinar sobre possibilidade e necessidade. Há uma parte da lógica que estuda o comportamento dos operadores modais -- “é necessário que” e “é possível que” --, chama-se *lógica modal*.

Vamos considerar a seguinte distinção: males justificados e males gratuitos. Um mal, M , é justificado se e somente se houver algum bem, B , cuja ocorrência seja dependente de M , e B suplante M ou M evite um mal ainda maior.

Por exemplo, suponha que você tem gangrena no dedinho do pé. Para evitar que o restante do seu pé e mesmo sua perna gangrene, o médico decide amputar seu dedinho. Perder o dedinho do pé é um mal, mas perder toda a perna seria um mal ainda pior. Assim, o mal causado pela perda do dedinho é justificado. Por outro lado, M é gratuito quando não for justificado, isto é, quando sua ocorrência não evitar um mal maior ou não propiciar um bem maior. Por exemplo, se o médico fosse preguiçoso e amputasse seu dedo mesmo que não estivesse gangrenado, mas apenas ferido, ele teria causado um mal desnecessário, injustificado, gratuito.

Alguns teístas pensam que é perfeitamente concebível que Deus permitisse a existência de certos males, para que alcancemos um bem maior. Esse bem maior é o livre-arbítrio. É supostamente o maior bem do qual os seres humanos desfrutam. E a possibilidade do mal é uma consequência lógica da posse do livre-arbítrio. Deus não poderia ter criado criaturas livres e ao mesmo tempo ter eliminado a possibilidade do mal.

Vejamos. Ter livre-arbítrio significa poder escolher livremente que curso de ação tomar. E escolher livremente uma ação x implica que eu poderia ter, se quisesse, escolhido y . E cada uma de nossas ações tem diversas consequências. Por exemplo, suponha que escolhi jogar uma pedra na vidraça da delegacia. Dentre as consequências dessa ação, estão a quebra da vidraça e a minha detenção. Assim, a possibilidade da minha detenção é implicada pela minha escolha de jogar a pedra. Ou considere a minha escolha de colocar fogo numa área de preservação ambiental. O alastramento do fogo pode matar vários animais. Como anteriormente, a possibilidade da morte de vários animais é implicada pela minha escolha de colocar fogo no mato. Extrapolando a partir desses exemplos, podemos ver que as nossas escolhas livres têm como consequência a possibilidade da ocorrência de vários males. Ao dar o livre-arbítrio para os humanos, Deus, portanto, cria a possibilidade do mal. A efetivação do mal só ocorre quando os seres humanos executam alguma ação que têm o mal como consequência.

Lembre-se que o central da defesa do livre-arbítrio é a existência de um mundo possível que Deus exista e permita o mal. Mas tudo o que temos até agora é a possibilidade do mal, decorrente de alguma escolha livre humana. Como ligar as duas coisas? A resposta está na falibilidade humana. Os seres humanos erram com frequência em seus juízos, o que os leva muitas vezes a

fazer escolhas cujas consequências são ruins. Dizer que um agente é falível significa que esse agente é suscetível ao erro; em termos modais, isso significa que para um determinado juízo que esse agente faz corretamente no mundo efetivo, há um conjunto de mundo possíveis em que ele erra. Some-se a isso o conjunto de mundos possíveis em que uma determinada ação produz vários males. Haverá, portanto, pelo menos um mundo possível em que uma escolha equivocada produzirá um mal.

Mas por que Deus não intervém nesse mundo e elimina o mal? Ora, se Deus elimina o mal de todos os mundos, isso implica que o mal não é uma consequência genuína do livre-arbítrio. Considere o seguinte cenário. Um pai deixa o filho sozinho em casa enquanto viaja. Ele diz ao filho que faça o que quiser, que o filho é livre para fazer o que quiser. Mas sem que o garoto soubesse, ele cuida para que ninguém na vizinhança lhe forneça drogas. Seja com quem fosse que ele procurasse, venderiam placebo a ele. Logo que o pai se ausenta, o garoto vai até à boca de fumo mais próxima e compra uma substância que ele acreditava ser alucinógena (mas que era na verdade placebo!). O garoto fez uma escolha livre, porém não sofreu as consequências dessa escolha. Qual é o valor do livre-arbítrio nesse caso?

Do ponto de vista do aprimoramento moral, nenhum. O garoto não pôde aprender com sua experiência. Além do mais, um pai que interfere desse modo nas escolhas do filho não permite que o filho se aprimore moralmente. E um pai que age assim é um pai defeituoso. Do mesmo modo, se Deus cuidasse para que o resultado das nossas escolhas livres não produzisse qualquer mal, não estaria permitindo que nos aprimorássemos moralmente. Por conseguinte, possuir livre-arbítrio não seria um bem. Ademais, uma vez que Deus não é um pai defeituoso, ele nos dará a capacidade de aprimoramento moral, através do uso do livre-arbítrio.

Em suma: Por ser sumamente bom, Deus nos dá o livre-arbítrio que, por sua vez, possibilita o mal. Além disso, o livre-arbítrio só tem valor, se pudermos aprender com os erros. E só podemos aprender com os erros, se Deus não interferir nas consequências de nossas escolhas livres. Portanto, para nos conceder o bem advindo do exercício do livre-arbítrio, Deus não pode eliminar o mal que porventura surja como consequência.

Resumamos. Começamos com o ataque de que as proposições *Deus existe* e *o mal existe* são inconsistentes entre si. Para derivar uma contração a

partir delas, tivemos de adicionar uma nova proposição, a de que Deus elimina o mal. Mas vimos que essa proposição não é necessariamente verdadeira, pois há pelo menos um mundo possível em que Deus existe e não elimina o mal -- a saber, aquele mundo em que a ação livre dos seres humanos provoca algum mal. Portanto, o argumento lógico do mal não pode ser bem sucedido.

4.1.2. O argumento evidenciário

Seu amigo não se rendeu à defesa do livre-arbítrio. Ele ainda acha que a existência do mal, porquanto não refute a existência de Deus, a torna improvável. Ou seja, ele pensa que a existência do mal fornece evidência indutiva a favor da existência de Deus.

O argumento evidenciário procura oferecer algum exemplo de mal gratuito. A melhor categoria de tais casos envolve o sofrimento de animais não-humanos, pois em geral esse tipo de sofrimento em nada contribui para o alcance de qualquer tipo de bem maior. O incêndio que assolou a Austrália no final de 2019 e início de 2020 matou mais de 1 bilhão de animais. Ficou famosa a foto de um filhote de canguru preso a uma cerca nas colinas de Adelaide, que acabou por morrer carbonizado. Talvez as chamas tenham atingido fatalmente esse filhote após ele ter desmaiado, caso que ele sofreria um pouco menos. Mas, ainda assim, o sofrimento causado pelo pânico durante a fuga, ou na tentativa de se soltar da cerca, enquanto o fogo se aproxima e a temperatura aumenta, não é desprezível. É difícil conceber algum bem que pudesse justificar a existência desse mal. Parece, portanto, ser um mal gratuito.

Podemos esquetizar o argumento evidenciário do seguinte modo:

1. Há males gratuitos.
2. Se há males gratuitos, então Deus não existe.
3. Logo, Deus não existe.

O argumento é válido, é um *modus ponens*. Talvez você esteja se perguntando qual é afinal a diferença entre os dois argumentos do mal; uma vez o argumento evidenciário também pretende mostrar que, se for verdade que males gratuitos existem, então não pode ser verdade que Deus existe. A diferença, contudo, reside no tipo de atitude que se tem frente às afirmações de

que há males e de que há males gratuitos. Que há males no mundo é algo que tanto teístas quanto não teístas concordam. Na verdade, é parte da doutrina cristã, por exemplo, que haja mal no mundo. O mesmo não se pode dizer da existência de males gratuitos; teístas e não teístas observam o mesmo evento de sofrimento, mas veem coisas diferentes: onde o não teísta vê um mal gratuito, o teísta tende a ver um mal justificado. Isso quer dizer que da observação de um evento de sofrimento não podemos deduzir a existência de um mal gratuito. A premissa 1 do argumento evidencial, portanto, depende de uma inferência indutiva:

1. M é um mal.
2. Parece não haver qualquer bem que justifique M.
3. Logo, (provavelmente) M é um mal gratuito.

Talvez você tenha já esboçado a seguinte reação. A incapacidade do não-teísta de conceber um bem que justifique o sofrimento do filhote de canguru se deve ao seu pouco conhecimento da doutrina religiosa, ou mesmo da sua falta de fé. No primeiro caso, um pouco de estudo cuidadoso de teologia e escritos religiosos talvez ajude. No segundo caso, só a prática religiosa e a estreita conexão com Deus poderia fazer com que um sujeito, mesmo não sendo capaz de conceber um bem maior que justifique aquele mal aparentemente gratuito, não fosse demovido da crença de que Deus sabe qual é esse bem maior.

4.1.3. Reações teístas

Há duas maneiras principais pelas quais você poderia responder a seu amigo. A primeira envolve oferecer um relato que explique por que Deus permite que haja mal no mundo; a tal explicação, damos o nome de “teodiceia”. Várias teodiceias têm como ponto central o livre-arbítrio. Por exemplo, vários cristãos pensam que uma relação genuína de amor com Deus envolve a aceitação livre do crente e o reconhecimento de que Deus é o criador e o ama. E como já vimos, não podemos fazer escolhas livres sem que seja possível a existência do mal. Como somos seres limitados cognitivamente e moralmente, é normal que as escolhas que fazemos durante a nossa vida tenham o mal como consequência. O

aprimoramento da alma humana envolve o erro. A nossa passagem na Terra é marcada pelo mal provocado por nossas escolhas livres; e lidar com o mal faz parte do processo do aperfeiçoamento de nossas almas e nos prepara para entrar numa relação amorosa com Deus.

Teodiceias não são fáceis de apresentar em um curto espaço; em geral, vários detalhes peculiares à teologia são necessários. Na falta de espaço, podemos explorar a segunda maneira. Ela diz respeito à nossa capacidade de especular sobre a mente de Deus. De acordo com essa resposta, as intenções de Deus nos são inescrutáveis -- *i.e.*, não podem ser compreendidas ou sondadas por nós. Sendo assim, um mal que nos parece gratuito porque não somos capazes de conceber um bem maior que o justifique certamente não parecerá gratuito para Deus. Dito de outro modo, a nossa incapacidade de sondar a mente de Deus nos proíbe (epistemicamente) de acreditar que um determinado mal aparentemente gratuito seja de fato gratuito. Essa posição se chama “teísmo cético”. O ceticismo do teísta aqui se restringe à possibilidade de saber ou acreditar justificadamente em proposições sobre o conteúdo da mente de Deus: suas vontades, intenções e crenças.

Lembre-se que o argumento evidenciário do mal é indutivo; ele depende do seguinte passo: *se me parece que p, então [é provável que] p*. Há alguma justificção para esse passo indutivo? Alguns filósofos pensam que uma evidência só é boa quando ela é uma indicação confiável da verdade. Por exemplo, a altura da coluna de mercúrio num termômetro é uma indicação confiável da temperatura do ambiente. Suponha que o termômetro marque 38,5°C; se não fosse verdade que a temperatura é 38,5°C (digamos que fosse 36°C), então o termômetro não marcaria 38,5°C (marcaria 36). Por outro lado, um termômetro que marcasse 38,5°C quando a temperatura fosse outra (36°C, por exemplo) não seria um indicador confiável.

Vejamos esse princípio em funcionamento na formação de crenças. Suponha que, numa operação policial, um recruta veja, à curta distância e sob boa iluminação, uma pessoa que se parece (no sentido de me ser apresentado aos sentidos) com o foragido Queiroz. Rapidamente ele reconhece que é de fato o Queiroz. A aparência física dessa pessoa (seus traços, sua careca, sua estaturaetc.) forneceu-lhe evidência para acreditar que era o Queiroz. Se não fosse o Queiroz (e sim uma outra pessoa, mesmo que parecida com ele), o recruta não teria acreditado que era o Queiroz. Portanto, a crença do recruta de

que é o Queiroz foi baseada numa indicação confiável; por conseguinte, ele tinha boa evidência.

Suponha agora que o Queiroz tenha um irmão gêmeo idêntico, o Nestor. Nas mesmas condições, o recruta olha para uma pessoa e vê alguém que parece o Queiroz. Mas poderia ser o Nestor. O recruta agora não tem boa evidência, uma vez que, se fosse o Nestor que estivesse em sua frente, ele acreditaria que é o Queiroz. Ou seja, ele tem a mesma resposta para inputs diferentes. Alguém parecer o Queiroz, nesse caso, não é uma indicação confiável de que é o Queiroz. É como o termômetro marcar 38,5°C quando a temperatura é de 36°C.

O teísta cético diz que a crença no mal gratuito não pode ser baseada em boa evidência; que a aparência de mal gratuito não é uma indicação confiável de mal gratuito. Considere novamente a morte do filhote de canguru. A aparência de mal gratuito se dá porque não conseguimos conceber um bem que justifique esse mal. Assim, de acordo com o nosso princípio epistêmico da boa evidência: se houvesse um bem maior que justificasse esse mal, então seríamos capazes de concebê-lo.

Vamos supor então que no mundo efetivo a morte do filhote é um mal gratuito. Imagine agora um mundo possível em quase tudo igual ao efetivo, com a única diferença de que há um bem que justifique o mal causado pela morte do canguru. Que bem é esse? Só Deus sabe. Os humanos são incapazes de conceber tal bem. Nesse mundo, embora haja justificação para a morte do filhote, continua parecendo aos humanos que esse mal é gratuito. Mas não é! E se é assim nesse mundo possível muito próximo do efetivo, por que não poderia ser assim também no mundo efetivo? Não dá para saber. Portanto, é falso que, se houvesse um bem maior, então seríamos capazes de concebê-lo. Por conseguinte, a aparência de mal gratuito não é uma boa evidência para pensarmos que a morte do filhote seja um mal gratuito.

Note que disso não se segue o que teísta saberá qual é esse bem. O ceticismo vale para os dois lados. O teísta se vale dessas considerações, lembre-se, para bloquear a inferência da aparência de mal gratuito para a existência do mal gratuito. Sendo assim, o defensor do argumento evidenciário não pode afirmar uma das premissas de seu argumento, designadamente a de que há males gratuitos no mundo. Vale notar, uma vez mais, que o teísta não está dizendo que não há males gratuitos; está dizendo apenas que não podemos

dizer se um mal é gratuito ou não. E isso é suficiente para seus propósitos.

4.2. O desacordo religioso

Sentindo-se bloqueado pelo teísmo cético, seu amigo não perdeu tempo e foi se informar sobre outras religiões. Ele descobriu que, em geral, quase todas as religiões têm o mesmo tipo de preocupação com o lugar que ocupamos no universo, com a relação entre criador e criatura, com a vida após a morte e assim por diante. Mas algumas diferem radicalmente sobre as respostas que fornecem a essas preocupações. Por exemplo, os cristãos acreditam que Deus é o criador do universo; os iorubás acreditam que é Olorum. Ambos não podem estar corretos. Ou Deus é o criador, ou Olorum. Embora não haja uma teologia natural iorubá, não é difícil imaginar que o mesmo tipo de argumentos possa ser construído a favor de Olorum. Sendo assim, teríamos conclusões contrárias apoiadas por razões igualmente boas.

Suponha que seu amigo lhe explica os argumentos iorubá. Imagine que esses argumentos têm premissas parecidas com os argumentos teístas. Assim, se você considera as suas premissas boas, também terá de considerar boas as do argumento iorubá. Por conseguinte, também terá de considerar racional a crença dos iorubás. Lembre-se que há aqui um desacordo: ou os teístas estão corretos ou os iorubás estão corretos, mas não ambos. Se você considera as duas crenças racionais, qual é atitude racional a ser tomada?

Seu amigo diz que você deve suspender o juízo. Vejamos. Frente a uma proposição qualquer (p), um sujeito (S) pode ter as seguintes atitudes doxásticas: a) acreditar que p , b) acreditar que não- p ou c) suspender o juízo. Se houver evidência suficiente a favor de p , então é racional que S acredite que p . Se houver evidência suficiente a favor de não- p , então é racional que S acredite que não- p . Mas se a evidência for insuficiente, é racional que S suspenda o juízo.

Há duas situações em que a evidência é considerada insuficiente. A primeira é quando a evidência aumenta muito pouco a probabilidade de a crença ser verdadeira. Por exemplo, você acredita que Carlos assassinou a própria esposa porque a arma do crime era uma pistola registrada em nome de Carlos. Essa evidência, ainda que dê um pequeno apoio à sua crença, não a torna provável o suficiente. Não seria racional da sua parte sustentar a crença de que Carlos é o assassino (o que não implica você ter de acreditar que Carlos é

inocente). A segunda é quando o sujeito adquire evidência igual para crença e sua negação. Suponha agora que além da arma pertencer a Carlos, são encontradas suas impressões digitais na arma, vestígios de pólvora na mão de Carlos e um recibo de uma caixa de munição comprada no dia anterior. Nesse caso, parece que a evidência se torna suficiente para acreditar que Carlos é o assassino. Não fosse por um novo conjunto de evidências. Os investigadores descobrem que Carlos é instrutor de tiro; que a sua esposa foi alvejada por uma munição diferente daquela comprada no dia anterior e que ela vinha sendo ameaçada pelo ex-marido. Essas evidências contrabalançam as evidências do primeiro conjunto. Assim, se é racional acreditar que Carlos é o assassino, também é racional acreditar que Carlos não é o assassino. Mas não pode ser racional acreditar que Carlos é e não é o assassino (acreditar em uma contradição não parece ser uma política racional!). Esse novo conjunto de evidências faz com que sua evidência total seja insuficiente para acreditar que Carlos é o assassino; e do mesmo modo insuficiente para acreditar que ele não é o assassino. Portanto, a atitude racional na falta de evidência suficiente é a suspensão da crença.

Ora, as mesmas considerações se aplicam ao desacordo religioso. O teísta tem boa evidência a favor de sua crença. E o iorubá tem evidência igualmente boa para a dele. Se você reconhece que ambos os conjuntos de evidência como razoáveis, então a sua evidência total é insuficiente para acreditar na crença teísta e também insuficiente para acreditar na crença iorubá. Portanto, é racional suspender a crença sobre a existência de Deus e de Olórum.

Começamos dizendo que o teísta e o iorubá não podem estar ambos corretos. Mas e se estiverem ambos errados? Pode ser que uma terceira crença esteja correta, por exemplo de alguns indígenas. Não poderia também haver aparentes boas razões para se aceitar tal crença? Obviamente que sim. Mas nesse caso, as aparentes boas razões do teísta e do iorubá anulariam a evidência para acreditar na crença indígena. E novamente poderíamos considerar uma nova crença religiosa, em desacordo com as três anteriores. Essa, por sua vez, seria anulada pelas evidências das crenças anteriores. E assim por diante.

Há uma maneira interessante de se responder a esse desafio; envolve a adoção de uma posição chamada de “pluralismo religioso”. A ideia básica do pluralismo é que não há uma religião correta – “A religião” – mas muitas. Há pelo

menos duas versões de pluralismo: o relativista e o construtivista.

Começemos com o relativismo. O relativista pensa que não há objetividade em questões religiosas. Ele pensa que o desacordo religioso mostra o quão dramática é a tentativa de estabelecer verdades religiosas. O conhecimento religioso é impossível. Não importa se existem ou não fatos transcendentais da religião. Se existirem, o desacordo mostra que é impossível conhecê-los; se não existirem, é óbvio que não podem ser conhecidos. Sendo assim, é permitido às diferentes culturas que tenham qualquer tipo de crença sobre a divindade. E virtualmente todas essas crenças estão em pé de igualdade; é um vale-tudo. Relatos religiosos são como histórias fictícias; cada história fictícia é única e não compete em realidade com outras histórias.

O pluralista construtivista pensa que há sim uma divindade transcendente. O cerne do pluralismo construtivista é a ideia de que há uma divindade transcendente e que as várias religiões “constroem” essa divindade (ou divindades), através de seus sistemas conceituais particulares. Assim como o relativista, o construtivista pensa ser impossível conhecer a realidade transcendente religiosa. Mas não porque há desacordo religioso e sim porque a própria divindade é em si incognoscível. Vejamos.

O construtivista pressupõe que o divino existe independente de nossas mentes; mesmo que ninguém concebesse ou pensasse no divino, ainda assim ele continuaria existindo. Contudo a mente humana só pode fazer representações do divino; todo acesso ao divino é mediado por representações humanas. Alguns representam a divindade como o deus teísta; outros como muitos deuses; há também quem a representa como uma força impessoal que rege o universo e assim por diante. De acordo com o construtivista, todos estão representando a mesma coisa. E para além dessas representações nada pode ser dito de substancial acerca da divindade.

Uma imagem pode nos ajudar. Pense num objeto microscópico. Ele não pode ser observado a olho nu; qualquer imagem que tenhamos dele terá de ser feita por um microscópio. Mas suponha que há vários microscópios diferentes e cada um deles produz uma imagem diferente do mesmo objeto. Além disso, cada uma dessas imagens é igualmente satisfatória para os cientistas que as produzem. Substitua o objeto microscópico pela divindade e os microscópios pelas diferentes religiões. Não há contato imediato com a divindade; sem uma tradição religiosa não se pode “ver” a divindade. E várias religiões apresentam

diferentes “imagens” da mesma coisa, cada uma delas igualmente satisfatória a seus praticantes.

É claro que alguém poderia protestar dizendo que nem todas as imagens geradas por microscópios diferentes são igualmente boas. Algumas representam melhor o objeto. Pode haver imagens complementares assim como imagens conflitantes. O trabalho dos cientistas é encontrar a melhor representação do objeto microscópico. O mesmo vale para a religião. Pode haver representações complementares do divino assim como representações conflitantes. O teólogo e o filósofo da religião deveriam encontrar a religião que melhor representa o divino, não dizer que todas as religiões são igualmente boas. Se você pensa assim, então você é um *exclusivista*. Mas se você concorda com a força argumento do desacordo, ser exclusivista lhe obrigará a suspender o juízo. Portanto, a melhor estratégia para evitar a acusação de irracionalidade é adotar o pluralismo.

5. A epistemologia da crença religiosa

Depois de dias refletindo sobre a sugestão de seu amigo de que a postura racional a ser tomada frente à crença religiosa é a suspensão da crença, você acaba por descobrir uma nova tradição de pensamento sobre a crença religiosa. Primeiro, ela não está focada nos argumentos da teologia natural, a sua abordagem diz respeito mais à racionalidade da crença religiosa. Segundo, ela fornece a chave para se entender por que parece razoável a uma grande quantidade de pessoas a adoção de uma postura agnóstica ou atea.

5.1. O desafio evidencialista

“Alegações extraordinárias exigem evidência extraordinária”. Você provavelmente já ouviu algum entusiasta repetindo esse mote a alguém que acredite em Óvnis, em atividades paranormais, etc. De certo modo, muitos agnósticos ou ateus pensam que isso se aplica à própria alegação de que Deus existe. E na falta de tais evidências, ou no caso delas serem insuficientes, não deveríamos acreditar que Deus existe. Insistir em manter essa crença tornaria a pessoa *irracional*. Ou seja, uma pessoa racional se guia pelas evidências.

Note que ao dizer que um sujeito é racional ou que a sua crença é

racional, isso envolve duas coisas: (i) o direito do sujeito a sustentar determinada crença e (ii) a justificação que o sujeito tem para essa crença. É a evidência que unifica as duas. Possuir evidência é uma condição necessária para o sujeito estar justificado a sustentar uma crença. A justificação é a propriedade que adicionada em grau suficiente à crença verdadeira a torna conhecimento. Mas adicionada à crença falsa (que o sujeito sustenta como verdadeira), dá o direito, ao sujeito, a sustentar tal crença. Não faz sentido culpar um sujeito por sustentar uma crença falsa, se ele fez o seu melhor, do ponto de vista evidencial, para sustentá-la. Por outro lado, sustentar uma crença verdadeira sem qualquer evidência é irresponsável.

O que importa sublinhar aqui é que, verdadeira ou falsa, a crença teísta só é racional no caso de haver boas evidências para sustentá-la. A tradição da teologia natural é evidencialista. Assim como os ateus e agnósticos também são. Todos eles pretendem guiar suas atitudes doxásticas de acordo com a disponibilidade de evidências. E vimos que seguir a política evidencialista pode acabar transformando um teísta exclusivista em um pluralista.

5.2 O argumento da paridade

Mas será que o evidencialista está certo? Será que as nossas crenças religiosas precisam estar baseadas em evidência? Alguns epistemólogos pensam que não. De acordo com eles, a crença na existência de Deus está entre um tipo de crenças cuja justificação não depende de evidência provinda de outras crenças, o que se chama comumente de “crenças básicas”. Exemplos típicos de crenças básicas são a crença de que o mundo externo existe, a crença de que as outras pessoas possuem mente, etc. A estratégia aqui será dizer que a crença teísta está em paridade epistêmica com as crenças de que há um mundo externo, de que há outras mentes, etc. Para usar uma metáfora, a crença teísta está no mesmo barco que essas crenças; se uma afundar, as outras afundam junto.

Vamos começar com a distinção entre crenças básicas e não-básicas. Considere a sua crença de que alguém revisou este texto que você está lendo agora. Qual é a justificação que você tem para essa crença? Decerto você recorreu a outras crenças. Por exemplo, a sua crença de que textos publicados são geralmente revistos por outras pessoas, que este texto é bastante complexo

para ter sido revisado por um algoritmo de revisão, etc. A partir dessas crenças você inferiu a crença de que este texto foi revisado por alguém. Essa é uma crença *inferencialmente justificada*; a justificção dela depende de outras crenças. Se você dependeu de inferência para justificar uma crença, então essa crença não pode ser básica; ela é não-básica. As crenças básicas são aquelas cuja justificção é *não-inferencial*. Além disso, as crenças básicas evitam que você caia em cadeias de justificções infinitas. Por exemplo, considere a sua crença de que textos publicados são geralmente revistos. A justificção que você tem para essa crença é inferencial ou não-inferencial. Se for inferencial, você precisará de uma nova crença, que, por sua vez, terá justificção inferencial ou não-inferencial, e assim por diante. Esse potencial regresso só é impedido quando atingimos uma crença básica.

As crenças perceptuais são o tipo mais comum de crença básica. A minha crença de que há um computador em minha frente no instante em que escrevo este texto, a minha crença de que tenho mãos, etc. são crenças que consideramos justificadas e cuja justificção é não-inferencial. As crenças básicas justificam outras crenças, mas não são justificadas por outras crenças.

Mas em que sentido as crenças religiosas são básicas? Tomemos como exemplo a crença do cristão de que seus pecados foram perdoados. Digamos que ele formou essa crença depois de uma profunda oração pedindo perdão. Muitos crentes formam crenças assim. E essas crenças não dependem de outras crenças. Não é que o cristão acredita que Deus perdoo depois de uma oração; antes, ele sente que foi perdoado. Ou considere o sentimento de maravilhamento que muitos têm diante de uma paisagem grandiosa. Há vários relatos de pessoas que formam a crença de que aquela beleza nos foi oferecida por Deus. Note, não é que o crente raciocine que Deus é o criador; muito pelo contrário, ele forma diretamente a crença de que o criador nos presenteou. Novamente, essa crença não depende de outras crenças, é uma crença básica.

Repare que as crenças básicas teístas são crenças cujo conteúdo em parte depende de Deus. Para que a minha crença de que Deus me perdoo seja verdadeira, é preciso também que seja verdade que Deus exista. Para que a minha crença de que *a beleza do cânion de Ribeirão das Bandeirinhas, na Serra do Cipó, é um presente de Deus* seja verdadeira, novamente tem de ser verdade que Deus existe. Do mesmo modo, a minha crença de que tenho mãos só é verdadeira, se for verdade que há objetos externos à minha mente que são as

minhas mãos. E a minha crença de que João está com dores só é verdadeira, se também for verdade que João tem uma mente.

Em geral, o teísta acredita que Deus existe do mesmo modo como qualquer pessoa acredita na existência do mundo externo e na existência de outras mentes. E na medida em que essas crenças estão justificadas, não estão em função do sujeito possuir evidência em favor delas.

Tomando as crenças perceptuais como exemplos paradigmáticos de crenças básicas, podemos isolar duas características das crenças básicas: (i) são justificadas não-inferencialmente e (ii) são formadas por mecanismos cognitivos confiáveis. E “confiável” aqui tem um significado preciso, é a tendência que o mecanismo tem de produzir mais crenças verdadeiras do que crenças falsas. Cuidado aqui. Ter uma tendência a formar mais crenças verdadeiras do que falsas não quer dizer que o mecanismo efetivamente forme mais crenças verdadeiras do que falsas. Por exemplo, a visão humana tende a produzir mais crenças verdadeiras do que falsas, quando a visão do sujeito está funcionando bem e em condições adequadas para esse funcionamento. A uma distância média, sob a luz do dia e não tendo problemas de visão, se olho para um objeto e formo a crença de que é um gambá – sendo de fato um gambá –, então sei que aquilo é um gambá. Se eu fosse míope, por exemplo, a chance de eu errar – e.g., acreditar que era um rato – se tornaria mais alta. Ou se a luz estivesse baixa, no crepúsculo, as chances de eu errar também passariam a ser maiores. Considerações similares podem ser feitas sobre os outros mecanismos cognitivos importantes: os outros sentidos, a memória e o raciocínio.

Mas e quanto à crença religiosa, haverá algum mecanismo cognitivo confiável responsável pela sua formação? As supostas crenças básicas teístas não são formadas pelos mecanismos perceptuais paradigmáticos. Se há alguma crença teísta básica, ela não pode brotar do nada no crente; é preciso que identifiquemos a fonte de tais crenças.

Alguns teístas pensam haver um senso do divino, ao qual costumam chamar *sensus divinitatis*. O *sensus divinitatis*, de acordo com eles, é confiável, uma vez que é um mecanismo cognitivo moldado por Deus e que foi dado aos seres humanos, para que tivessem exatamente a função de produzir crenças religiosas verdadeiras. Aquelas crenças sobre a grandeza da natureza, sobre ter sido perdoado, etc. são crenças exatamente do tipo formado por muitos teístas.

O mérito da epistemologia reformada reside em proteger a crença teísta

do ataque de irracionalidade. Primeiro, ao considerar a crença teísta como básica, ela evita o desafio evidencialista. Segundo, e mais interessante, ela inverte o ônus para o opositor do teísmo. Na medida em que a crença teísta é racional, a acusação de irracionalidade passa a depender da negação da existência de Deus, uma vez que negar a existência do *sensus divinitatis* é rejeitar o relato teísta. Ou seja, a epistemologia reformada obriga o opositor do teísmo a primeiro estabelecer que a crença teísta é falsa.

Uma dúvida natural aqui é: por que nem todas as pessoas formam crenças através do *sensus divinitatis*? Afinal, se Deus projetou esse mecanismo cognitivo para o ser humano, por que há ateus e agnósticos?

A resposta dificilmente convencerá o agnóstico: o senso do divino é afetado pelo pecado e suas consequências. Assim como os mecanismos perceptuais paradigmáticos estão sujeitos ao mal funcionamento, também o *sensus divinitatis* está. Por exemplo, o mal funcionamento da visão pode levar à cegueira. Do mesmo modo, o mau funcionamento do senso do divino está sujeito à “cegueira”. Restabelecer o senso do divino pode envolver a prática religiosa, o arrependimento, a restituição da alma, etc.

O teísta, contudo, pode enfatizar o seguinte. Não há outro modo de compreensão do senso do divino a não ser a experiência em primeira pessoa ou o testemunho. É como as experiências de quase-morte. Ou você acredita nelas, porque já teve alguma experiência do tipo, ou porque aceita o relato de pessoas que disseram ter tido. Não há como detectar experiências de quase morte num experimento controlado. Se o agnóstico pensa ser racional a aceitação de relatos de quase-morte, por que não pensaria também serem os relatos das experiências produzidas pelo senso do divino?

Há, porém, uma dificuldade maior. Considere uma comunidade de pastafarianistas, que acreditam que somos a criação do Monstro do Espaguete Voador. Suponha que alguns teólogos e filósofos pastafarianistas tenham percebido a existência de crenças pastafarianistas básicas naquela comunidade. Eles consultam as escrituras sagradas do Pastafarianismo e descobrem a existência de um mecanismo cognitivo sobrenatural, o *sensus espagueticus*, responsável pela produção de crenças básicas pastafarianistas. Oficializa-se então a epistemologia religiosa pastafarianista, que defende que as crenças pastafarianistas são racionais. Ora, se as crenças dos cristãos são racionais devido ao *sensus divinitatis*, por que as crenças dos pastafarianistas não seriam

racionais devido ao *sensusespagueticus*?

O teísta aqui se vê novamente no problema do desacordo. Se aceita que a crença pastafarianista é racional terá de ser pluralista. Por outro lado, se quiser afastar como implausível a epistemologia pastafarianista, terá de argumentar a favor do teísmo. Pois, assim como a epistemologia reformada pressupõe a existência de Deus, a epistemologia pastafarianista supõe a existência do Monstro do Espaguete Voador. E Deus não pode coexistir com o Monstro do Espaguete Voador. Assim, o teísta só conseguirá afastar a epistemologia pastafarianista na medida em que conseguir mostrar que o teísmo é mais plausível que o pastafarianismo. No fim das contas, o epistemólogo reformado se vê diante do mesmo tipo de desafio que ele próprio impusera ao agnóstico: argumentar contra a verdade da crença religiosa (nesse caso, contra a crença pastafarianista).

6. Naturalismo e sobrenaturalismo

As discussões entre seu amigo e você têm sido proveitosas. Embora tenha considerado os argumentos da tradição reformada, e seja bastante tolerante quanto a crença religiosa, ele ainda sente haver melhores razões para a descrença do que para a crença. Ele pensa que deveríamos ter atitudes mais próximas às dos cientistas. A instituição científica funciona muito bem sem Deus. Ela investiga questões pertencentes ao *mundo natural*. O sucesso dos métodos científicos no tratamento das mais diversas questões parece não deixar dúvida de que nada há na realidade que não seja natural. A crença teísta tem como objeto um ser sobrenatural. Deus não é objeto de uma ciência como a física ou a biologia. Na medida em que Deus faz parte do mundo sobrenatural, e não temos boas razões para pensar que haja qualquer realidade sobrenatural, a existência de Deus não deveria ser levada a sério do ponto de vista intelectual. O fenômeno religioso certamente merece ser estudado, mas do ponto de vista do mundo natural, como o estudam a sociologia, a história, a psicologia, etc.

6.1. Naturalismo

O nível 0 do naturalismo -- *i.e.*, aquilo que todos os naturalistas têm em comum -- é a ideia de que não há espaço para entidades que não sejam do

mundo natural na teorização científica e demais atividades cognitivas humanas. O nível 1, por assim dizer, sustenta que não deveríamos nos comprometer com entidades que não sejam do mundo natural porque a metodologia científica bem sucedida não está comprometida com tais entidades. O nível 2 diz que o sucesso da metodologia científica nos dá as melhores razões para acreditar que tudo o que há existe no mundo natural e em nenhum outro. Vamos chamar o nível 1 de “naturalismo metodológico” e o nível 2 de “naturalismo ontológico”. O naturalismo ontológico é mais forte que o metodológico. O primeiro implica que temos as melhores razões para acreditar que não há entidades sobrenaturais. O último não; implica que devemos nos comprometer apenas com o tipo de entidades que a nossa melhor ciência se compromete.

Ainda não está claro, porém, o que significa dizer que uma entidade “pertence ao mundo natural”. Isso talvez provoque alguma confusão, uma vez que na linguagem do dia-a-dia costumamos usar “natural” como oposto a “artificial”. Somos tentados a dizer que a casa do João-de-Barro é natural, ao passo que os conjuntos habitacionais da cidade grande são artificiais. A casa do João-de-Barro foi feita a partir de elementos extraídos diretamente da natureza, ao passo que um prédio é feito a partir de coisas que não são encontradas *in natura*. Na linguagem da filosofia, porém, nem sempre os termos do nosso vocabulário do dia-a-dia mantêm seus significados correntes. O conceito que estamos tentando captar aqui engloba tanto a casinha do João-de-Barro quanto modernas edificações. Na verdade, ele engloba de átomos (e seus constituintes) até sociedades humanas. Assim, diremos que algo é natural quando é, em última instância, formado por átomos (e, por conseguinte, seus constituintes). Estamos nos referindo à natureza em seu nível mais básico. Dado que os átomos são partículas físicas, o naturalista costuma dizer que tudo o que existe é, em última instância, físico. Sendo assim, os naturalistas se comprometem com o *fisicismo* – a tese segundo a qual tudo o que existe é físico.

O sobrenaturalismo, por sua vez, é a negação do naturalismo. É a tese segundo a qual além das coisas físicas, existem também coisas não físicas. Mas o que seriam essas coisas? Já sabemos que o teísta é sobrenaturalista, mas será que todo sobrenaturalista é teísta? A resposta é “não”. Um ateu pode ser sobrenaturalista. Mas isso depende da maneira como entendemos o conceito de algo ser sobrenatural. Novamente, na maneira de falar do dia-a-dia, somos tentados a considerar como sobrenatural apenas supostos fenômenos

relacionados ao além-mundo: milagres, comunicação de espíritos com médiuns etc. Mas o termo “sobrenatural” ganha um sentido técnico na filosofia. Se algo natural é algo físico, então algo sobrenatural é algo não físico. Geralmente os objetos físicos possuem localização no espaço. O computador que uso para escrever possui uma localização espacial determinada; além disso, ele não pode ocupar mais de uma localização ao mesmo tempo. E isso vale para qualquer objeto físico. Deus, por exemplo, não está sujeito às leis da física e ao espaço; ele não tem localização espacial. Assim, uma condição necessária para algo ser sobrenatural é não ter localização espacial. Ora, entidades matemáticas –e.g., números, conjuntos, proposições etc. – e entidades abstratas em geral não têm localização espacial. Poderemos dizer que o número 2 é uma entidade sobrenatural? Isto é, não ter localização espacial é também suficiente para que algo seja sobrenatural? Ou será que entidades sobrenaturais e entidades abstratas são categorias diferentes de entidades não físicas? Não precisamos resolver isso aqui (fica como exercício a reflexão sobre as entidades não físicas!). Para os nossos propósitos, tudo o que precisamos é da seguinte afirmação: *se o naturalismo for verdadeiro, então Deus não existe.*

Quadro 5: O projeto naturalista

Alguns filósofos naturalistas possuem um projeto filosófico ambicioso. Eles pensam poder responder a todos os grandes problemas filosóficos de um ponto de vista naturalista. Se tudo o que existe é físico, então a mente, embora pareça ser uma substância não física, tem de ser explicada em termos físicos. Julgamentos morais e estéticos, se são passíveis de ser verdadeiros ou falsos, dependem de fatos naturais. Se não, precisam de explicações compatíveis com o naturalismo. Objetos da matemática e da lógica, na medida em que desempenham papéis importantes na teorização científica, precisam também de uma explicação em termos naturais; ou no mínimo uma explicação que não recorra a entidades abstratas.

Os naturalistas costumam rejeitar uma metodologia completamente *a priori* para a filosofia. Eles não pensam que todas as questões filosóficas possam ser resolvidas sem o apelo à experiência, puramente *a priori*. Ao invés, veem-na num contínuo em que num extremo estão as ciências e no outro está a filosofia. Isso não quer dizer que os naturalistas se vejam como cientistas; quer dizer apenas que os filósofos deveriam prestar mais atenção às ciências empíricas. Descobertas científicas podem modificar teorias filosóficas, assim como concepções filosóficas podem influenciar a pesquisa científica.

6.2. O argumento evolucionista contra o naturalismo

Muitos teístas se sentem intimidados com o sucesso da ciência. Há muita propaganda ideológica envolvendo a tal da guerra (já se percebe pelo vocabulário bélico) entre ciência e religião. Há por um lado aqueles que pensam que ciência e religião são incompatíveis. São os incompatibilistas radicais que travam essa guerra. Mas não precisa ser assim, há os incompatibilistas moderados, que pensam que ciência e religião tratam de esferas diferentes. Por outro lado, há aqueles que acham a ciência e a religião perfeitamente compatíveis entre si. E dentre esses últimos, há ainda quem pense que são complementares. Desse ponto de vista, não é a ciência que conflita com a religião, mas sim o naturalismo. E o naturalismo é uma doutrina filosófica, não científica. É apenas uma contingência histórica que ciência e naturalismo tenham se confundido; não há qualquer ligação conceitual forte que obrigue a ciência ser naturalista, pensa o teísta.

Um argumento para mostrar que ciência e religião podem ser complementares parte de uma constatação compartilhada por teístas e naturalistas: o ser humano possui conhecimento. O conhecimento humano é o produto do funcionamento das nossas capacidades cognitivas. O naturalista, além disso, confia no relato da biologia evolutiva de que as nossas capacidades cognitivas são o produto de milhões de anos de ação da seleção natural sobre o cérebro dos nossos ancestrais. Vamos chamar isso de “evolucionismo naturalista”. O teísta concorda com o evolucionismo, mas não com o naturalismo. Com um pouco de reflexão sobre o padrão de explicação evolutiva, argumenta o teísta, podemos ver que o evolucionismo naturalista é falso. Por outro lado, o teísta apontará que a adoção do sobrenaturalismo teísta está perfeitamente de acordo com os resultados da história evolutiva humana.

Vamos explicitar o argumento:

1. Se o evolucionismo naturalista é verdadeiro, então o ser humano não tem conhecimento.
2. Mas o ser humano tem conhecimento.
3. Logo, o evolucionismo naturalista é falso.

A premissa 2, como já dito, é consenso. Mas é premissa 1 é o oposto do

que defende o naturalista. O teísta sabe disso e não espera que o naturalista a aceite. O que ele tenta fazer é mostrar que o naturalista está racionalmente obrigado a aceitá-la.

Vamos começar pelo conhecimento. Uma concepção bastante aceita sobre o conhecimento diz que conhecimento é crença verdadeira produzida por um mecanismo cognitivo confiável.

A ideia do naturalista é que há um relato evolutivo que explica por que os nossos mecanismos cognitivos são confiáveis –i.e., tendem a produzir mais crenças verdadeiras do que falsas. Do ponto evolutivo, uma espécie tende a manter certas características fenotípicas quando essas características ajudam os membros da espécie a sobreviver e se reproduzir, e, por conseguinte, a transmitir seus genes. Sendo assim, seja qual for o relato evolutivo, espera-se que ele mostre que ter mecanismos cognitivos confiáveis desempenhou um papel importante em aumentar as chances dos membros da nossa espécie em sobreviver e se reproduzir.

O teísta apresenta um experimento mental interessante a fim de mostrar que, do ponto de vista probabilístico, deveríamos esperar que a espécie *homo sapiens* tenha evoluído com uma tendência para acreditar mais em falsidades do que em verdades, isto é, que não evoluímos para obter conhecimento.

Eis o experimento mental. Considere o ancestral da nossa espécie. Ele vive na savana, um ambiente aberto e com poucas árvores. Há por ali um grande predador, tão voraz que é capaz de devorar um bando inteiro de primatas em questão de horas. Para sobreviver nesse ambiente, nossos ancestrais tinham de ter uma boa capacidade de fuga e proteção contra os ataques desse predador. É intuitivo pensarmos que esse primata teria uma grande vantagem se formasse um conjunto de crenças verdadeiras sobre o entorno e sobre seus predadores. Nesse cenário, podemos especular que a tendência é que os indivíduos que formam mais crenças verdadeiras sobrevivem, acasalam e passam adiante os genes que determinam a formação de mecanismos cognitivos confiáveis.

Considere, porém, que o nosso primata não forma crenças verdadeiras sobre o predador, mas sim crenças falsas. O tipo de mecanismo cognitivo que ele tem faz com que ele erre muito mais do que acerte quando pensa no seu predador. Acontece que essas crenças falsas fazem com que ele consiga se manter longe do predador e por isso sobreviver e se reproduzir. Nesse caso, o nosso primata passa adiante os genes para mecanismos cognitivos infiáveis (i.e.

não-confiáveis), de modo que seus descendentes também o possuirão.

Note que temos dois cenários possíveis. No primeiro, o nosso ancestral é selecionado porque tem crenças verdadeiras; no segundo, é selecionado porque tem crenças falsas. Na verdade, o que desempenha o papel central na sobrevivência do nosso primata é o seu comportamento. As suas crenças (junto dos seus desejos de permanecer vivo, não sofrer, etc.) produzem certo comportamento. Ou seja, o tipo de comportamento que permitiu o nosso ancestral sobreviver pode ter sido causado tanto por crenças verdadeiras quanto por crenças falsas. Isso quer dizer que a evolução é “cega” quanto à verdade. Do ponto de vista evolutivo, não importa se um indivíduo só forme crenças falsas, o que importa é que ele sobreviva e se reproduza. Mas se foi causado por crenças falsas, então é de se esperar que seus descendentes tenham evoluído para adquirir mais crenças falsas do que verdadeiras.

E aqui é que entra a parte principal do argumento que o teísta tem a favor da premissa 1. Vamos considerar aquelas crenças que causam o comportamento adaptativo. Para cada crença verdadeira que possa produzir um comportamento adaptativo, há muitas crenças falsas que podem produzir esse mesmo comportamento. Por exemplo, suponha que nosso primata, ao fugir do predador, tenha a crença verdadeira de que o predador é perigoso. Mas ele também poderia estar fugindo do predador porque acredita que está brincando de esconde-esconde, ou porque acredita que o predador é invejoso, e assim por diante. O teísta quer com isso mostrar que é mais provável que a evolução tenha trabalhado para produzir indivíduos com mecanismos cognitivos infiáveis (o que inclui a nossa espécie!) do que com mecanismos confiáveis. Uma analogia talvez nos ajude a compreender. Suponha que coloquemos 100 bolinhas num globo. 99 delas são pretas e uma vermelha. Depois de sacudir bem o globo e tirar uma bolinha, o que é mais provável ter acontecido: ter saído uma bolinha preta ou uma bolinha vermelha? Obviamente que é muito mais provável que tenha saído uma bolinha preta; as chances da bolinha vermelha ter saído são muitíssimo pequenas, uma em cem (1/100). Pense agora que a bolinha vermelha representa o par crença verdadeira/comportamento adaptativo e cada uma das bolinhas pretas representam um par diferente de crença falsa/comportamento adaptativo. O globo representa a seleção natural. Girar o globo é como se o processo de seleção natural estivesse acontecendo. O que é mais provável que aconteça agora: uma bolinha vermelha (crença verdadeira/comportamento) ou uma

bolinha preta (crença falsa/comportamento). Se você concorda com o resultado do primeiro caso, também tem de concordar com esse. O mesmo vale para o naturalista. A ideia dele de que a evolução, um processo natural sem qualquer interferência sobrenatural, é responsável por todas as características de uma espécie, incluindo seu aparato cognitivo, tem a consequência de que os seres humanos possuem um aparato cognitivo infiel. E como o conhecimento exige um aparato confiável, segue-se que os seres humanos não têm conhecimento.

O teísta, por outro lado, não enfrenta esse problema. Do ponto de vista teísta, Deus criou o mundo como um grande programador. Pensou em cada uma das leis da natureza e previu com grandes chances de acerto como estaria o planeta Terra depois do aparecimento do primeiro organismo vivo. E há duas possibilidades quanto à evolução humana: Deus pode ter intervindo no curso da história natural da evolução para criar o *homo sapiens* e sua notável capacidade para o conhecimento, ou pode ter programado a evolução, já durante a criação, para que ela conduzisse ao “homem que sabe”. Assim, a hipótese do evolucionismo sobrenaturalista ou teísta está de acordo com o dado inegável de que temos conhecimento, ao contrário do evolucionismo naturalista.

Se você prestou atenção, deve ter percebido que o evolucionismo naturalista nada mais é do que uma hipótese de desígnio inteligente: Deus projeta certa característica do mundo, nesse caso, o aparato cognitivo humano. Assim, na medida em que o teísta desbanca a hipótese naturalista e propõe a sua como melhor explicação, ele está oferecendo uma versão particular do argumento do desígnio.

Mas será que o seu amigo se convenceu do argumento? Será que um experimento mental sobre a evolução humana é capaz de estabelecer a premissa 1? Afinal, ela diz respeito a um capítulo da história da evolução biológica, e não podemos fazer história do que quer que seja sentados na poltrona, criando experimentos mentais. Podemos, no máximo, ler a história já escrita e tentar extrair suas consequências filosóficas. Mas terá o teísta lido essa história de maneira satisfatória? Podemos nos perguntar se, lendo tal história, conseguiremos estabelecer uma conexão conceitual forte entre verdade e sobrevivência. Pensar sobre esse argumento é um ótimo pretexto para irmos atrás de nova informação e aprender ciência.

Talvez nunca sejamos capazes de resolver a questão da existência de Deus; mas pensar sobre ela nos ajuda a expandir a compreensão da diversidade

de opiniões, dos limites do entendimento humano, da explicação científica e das teorias filosóficas. Pensar sobre a existência de Deus nos ajuda a entender a beleza e complexidade do imenso universo que habitamos. Permita-me terminar com as palavras de um dos maiores filósofos do século XX, Bertrand Russell, às quais eu jamais produziria iguais:

O homem sem rudimentos de filosofia passa pela vida preso a preconceitos derivados do senso comum, a crenças costumeiras da sua época ou da sua nação, e a convicções que cresceram na sua mente sem a cooperação ou consentimento da sua razão deliberativa. Para tal homem o mundo tende a tornar-se definitivo, finito, óbvio; os objetos comuns não levantam questões, e as possibilidades incomuns são rejeitadas com desdém. Pelo contrário, mal começamos a filosofar, descobrimos [...] que mesmo as coisas mais cotidianas levam a problemas aos quais só se podem dar respostas muito incompletas. A filosofia, apesar de não poder dizer-nos com certeza qual é a resposta verdadeira às dúvidas que levanta, é capaz de seguir muitas possibilidades que alargam os nossos pensamentos e os libertam da tirania do costume. Assim, apesar de diminuir a nossa sensação de certeza quanto ao que as coisas são, aumenta em muito o nosso conhecimento quanto ao que podem ser; remove o dogmatismo algo arrogante de quem nunca viajou pela região da dúvida libertadora, e mantém vivo o nosso sentido de admiração ao mostrar coisas comuns a uma luz incomum. (Russell, 1912)¹⁴⁵

¹⁴⁵ Gostaria de agradecer à Associação Brasileira de Filosofia da Religião (ABFR) cujo financiamento à minha pesquisa atual acabou por propiciar-me o tempo necessário para a produção deste artigo.

Questões para revisão

1. Quais são as propriedades de Deus? Explique cada uma delas.
2. Explique por que o argumento ontológico é um argumento *a priori*.
3. Explique a diferença entre argumentos *a priori* e *a posteriori*.
4. Por que um regresso infinito de causas não explica a existência do universo?
5. Construa exemplos de argumentos abduativos.
6. Qual a diferença entre o argumento lógico e o argumento evidencial do mal?
7. De que modo o livre-arbítrio justifica a ocorrência de males no mundo?
8. Como o desacordo religioso leva à suspensão do juízo?
9. Por que a crença religiosa pode ser considerada básica?
10. Explique como o argumento evolucionista contra o naturalismo pode ser transformado numa versão de argumento do designio?

Questões para discussão

1. Você concorda que Deus é por natureza onipotente, onisciente, onipresente, sumamente bom e um existente necessário? Faz sentido conceber Deus como não tendo algumas dessas propriedades?
2. É possível provar a existência de algo sem o auxílio da experiência? Se for, que tipos de coisas se pode provar que existe apenas pelo pensamento?
3. Suponha que Deus criou o mundo para que os humanos tivessem a oportunidade de aperfeiçoar suas almas através do exercício do livre-arbítrio e, no final, poder escolher livremente entrar numa relação de amor com Deus. O quão plausível lhe parece essa história? Justifique (de preferência sem pressupor o naturalismo).
4. O *sensus divinitatis* é um suposto mecanismo cognitivo produtor de crenças religiosas (cristãs) não-inferenciais. É consenso que os mecanismos perceptuais são boas fontes de crenças não-inferenciais sobre o mundo externo. Compare o *sensus divinitatis* como os mecanismos perceptuais padrão (visão, audição etc.) e discuta a plausibilidade de tal mecanismo.
5. Você concorda que não há um conflito real entre religião e ciência, mas sim entre religião e naturalismo? É possível acreditar em Deus e na ciência sem ao mesmo tempo ser irracional?

Sugestões de leitura

Leituras gerais

Há vários capítulos curtos sobre filosofia da religião em introduções gerais à filosofia. Você encontrará capítulos interessantes em: *Pense* (1999) de Simon Blackburn, *Problemas da Filosofia* (2005) de James Rachels e *O Básico da Filosofia* (1995) de Nigel Warburton. Os capítulos “Filosofia da Religião” de Charles Taliaferro no *Compêndio de Filosofia*, org. Bunnin & Tsui-James, e “Filosofia da Religião” de Agnaldo Cuoco Portugal em *Filosofia: uma introdução por disciplinas* (2012), org. Pedro Galvão são também proveitosos. Veja também o artigo “Existência de Deus” (2014), de Agnaldo Cuoco Portugal, no *Compêndio em Linha de Problemas de Filosofia Analítica*.

Vários argumentos apresentados aqui são apresentados de maneira esquemática em *Os 100 argumentos mais importantes da filosofia ocidental* (2011), organizado por Michael Bruce e Steven Barbone.

Para uma introdução mais abrangente, veja *Introdução à Filosofia da Religião* (2004) de William Rowe. Veja também *Um mundo sem Deus: ensaios sobre o ateísmo* (2007), org. Michael Martin.

Para conhecer um pouco melhor a tradição teísta você pode consultar as várias versões disponíveis da Bíblia. Em *Deus: uma biografia* (1995), de Jack Miles, você encontrará a bíblia de uma perspectiva literária. Em *Uma história de Deus* (1993), de Karen Armstrong, você encontrará uma história das principais concepções do deus teísta. Vale a pena também conhecer alguma discussão teológica, a qual você poderá encontrar em *Teologia: os fundamentos*, de Alister McGrath.

Leituras específicas

A formulação clássica do argumento ontológico encontra-se no *Proslogion* (1078) de Anselmo de Cantuária. Uma versão diferente pode ser encontrada na terceira meditação das *Meditações Metafísicas* (1641) de René Descartes. Uma versão moderna do argumento encontra-se em *Deus, a liberdade e o mal* (1974) de Alvin Plantinga. Para um menu mais completo dos diferentes argumentos ontológicos, veja o verbete da SEP “Ontological

Arguments”.

A versão do argumento cosmológico baseada no PRS foi primeiramente formulada por Samuel Clarke e Gottfried Leibniz. Uma versão acessível encontra-se em Rowe (2004). Uma versão anterior, o argumento tomista da causa primeira encontra-se na Suma Teológica (1485) de Tomás de Aquino. William Lane Craig e J. P. Moreland defendem uma versão diferente do argumento da causa primeira, o argumento *kalam*, em *Philosophical Foundations for a Christian Worldview* (2003).

Uma formulação usual do argumento do desígnio se encontra nos *Diálogos sobre a Religião Natural* (1779) de David Hume, cap. X, e em *Natural Theology* (1802) de William Paley. Um primeiro conjunto poderoso de críticas ao argumento do desígnio é apresentado no *Diálogos* de Hume. Um exame do argumento através das lentes da biologia evolutiva é magistralmente conduzido em *O Relojoeiro Cego* (1986), de Richard Dawkins. O argumento da complexidade irreduzível é desenvolvido em *A caixa preta de Darwin* (1996), do bioquímico Michael Behe. Uma versão do argumento do ajuste fino pode ser encontrada na monumental *A existência de Deus* (2004), de Richard Swinburne.

O argumento lógico do mal encontra-se em “O mal e a onipotência” (1955) de John Mackie. A defesa do livre-arbitrio pode ser lida em *Deus, o mal e a liberdade* (1974), de Plantinga. O argumento evidenciário é apresentado por William Rowe em “O problema do mal e algumas variedades de ateísmo” (1979) e o teísmo cético defendido em “O obstáculo humano aos argumentos indiciários do sofrimento: evitar os males da aparência” (1984) de Stephen Wykstra. Uma perspectiva interessante sobre o mal gratuito é desenvolvida em *O Problema do Mal* (3003) de Peter van Inwagen. John Hick discute algumas teodiceias e defende a sua, da edificação da alma, em seu *O mal e o Deus do amor* (2010). Os artigos de Mackie, Rowe, Wykstra, dentre vários outros artigos seminais sobre o problema do mal encontram-se na coletânea *O problema do mal: uma antologia de textos* (2013), organizada por S. R. N. Miranda.

Uma boa introdução ao problema do desacordo pode ser encontrada no verbete da IEP, “Religious Disagreement”. Uma defesa do pluralismo é desenvolvida no sofisticado *Uma interpretação da Religião* (2004) e em *Teologia Cristã e Pluralismo Religioso* (1995), ambos de John Hick.

O ataque evidencialista clássico foi oferecido por William Clifford em “A ética da crença” (1876). O principal defensor da epistemologia reformada é Alvin

Plantinga; seus argumentos aparecem em *Crença Cristã Garantida* (2000). Uma versão resumida é apresentada em *Conhecimento e crença cristã* (2015). Outro representante da epistemologia reformada é William Alston. Ele defende a racionalidade das experiências religiosas em *Percebendo Deus* (1991). O artigo de Clifford pode ser encontrado em *A Ética da Crença* (2010), organizado por Desidério Murcho.

Visões gerais do projeto naturalista podem ser encontradas em *Naturalismo* (2008) de Jack Ritchie e *Nosso Lugar no Universo* (1989) de J. J. C. Smart. Uma discussão sobre visões de mundo encontra-se em *Universo ao lado*, de James Sire (2009). Uma versão inicial do argumento evolucionista contra o naturalismo de Plantinga foi originalmente publicado em *Warrant and Proper Function* (1993). Uma versão resumida encontra-se em *Naturalism Defeated* (2002), org. James Beilby. Em um livro recente, *Ciência, Religião e Naturalismo: onde está o conflito* (2011), Plantinga discute em mais profundidade o debate entre ciência e religião.

Para uma perspectiva diferente veja *Pilares do Tempo* (1999), de Stephen Jay Gould. O estudo inicial da religião de um ponto de vista naturalista encontra-se em *História da Religião* (1757); uma defesa forte da religião como fenômeno natural é oferecida por Daniel Dennett em *Quebrando o encanto* (2006). Ernst Mayr apresenta uma introdução acessível à teoria da evolução em *O que é a evolução* (2001). A exploração da teoria da evolução para além da biologia é assunto de *A Perigosa Ideia de Darwin* (1995) de Daniel Dennett.

Referências

- ALSTON, W. (1991) *Percebendo Deus*. Trad. Agnaldo C. Portugal. Carisma 2020.
- ANSELMO, S. (1078) *Proslógio*. Trad. Angêlo Ricci. In *Os Pensadores*. Abril Cultural, 1988.
- ARMSTRONG, K. (1993) *Uma história de Deus*. Trad. Marcos Santarrita. Companhia das Letras, 2012.
- AQUINO, T. (1485) "As cinco vias". Trad. L. H. Marques Segundo. URL = <https://epistemicufop.wordpress.com/2018/09/15/as-cinco-vias-por-tomas-de-aquino/>.
- BEHE, M. (1996) *A caixa preta de Darwin: o desafio da bioquímica à teoria da evolução*. Zahar, 1997.

- BLACKBURN, S. (1999) *Pense*. Gradiva, 2001.
- BRUCE, M. & BARBONE, S. (2011) *Os 100 argumentos mais importantes da filosofia ocidental*. Trad. Ana Lúcia da Rocha Franco. Cultrix, 2013.
- CRAIG, W. L., & MORELAND, J. P. (2003) *Philosophical Foundations for a Christian Worldview*. InterVarsity Press.
- DAWKINS, R. (1986) *O relojoeiro cego: a teoria da evolução contra o designio divino*. Trad. Laura Teixeira Mota. Companhia das Letras, 2001.
- DENNETT, D. (1995) *A perigosa ideia de Darwin: a evolução e os significados da vida*. Tradução de Talita M. Rodrigues. Rocco, 1998.
- DENNETT, D. (2006) *Quebrando o encanto: a religião como fenômeno natural*. Tradução de Helena Londres. Globo, 2012.
- DESCARTES, R. (1641) *Meditações Metafísicas*. Trad. J. Ginsburg e Bento Prado Jr. Os Pensadores. Abril Cultural, 1973.
- GALVÃO, P. (org.) (2011) *Filosofia: uma introdução por disciplinas*. Edições 70.
- GOULD, S. J. (1999) *Pilares do tempo: ciência e religião na plenitude da vida*. Trad. F. Rangel. Rocco, 2002.
- HICK, J. (1995) *Teologia cristã e pluralismo religioso*. Trad. Luiz Henrique Dreher. Attar Editorial, 2005.
- HICK, J. (2004) *Uma interpretação da religião*. Trad. Agnaldo C. Portugal. Vozes, 2018.
- HICK, J. (2010) *O mal e o Deus do amor*. Trad. Sérgio Miranda. Editora UnB, 2018.
- INWAGEN, P. v. (2003) *O problema do mal*. Trad. Sérgio Miranda. Editora UnB, 2018.
- HUME, D. (1757) *História natural da religião*. Trad. Jaimir Conte. São Paulo: Unesp, 2005.
- HUME, D. (1779) *Diálogos sobre a religião natural*. Trad. Bruna Frascolla. Salvador: EDUFBA, 2016. <https://doi.org/10.7476/9788523218676>.
- MARTIN, M. (org.) (2007) *Um mundo sem Deus: ensaios sobre o ateísmo*. Edições 70, 2010.
- MAYR, E. (2001) *O que é a evolução*. Trad. R. S. de Biase e S. C. de Biase. Rocco, 2009.
- MCGRATH, A. (2008) *Teologia: os fundamentos*. Trad. Joshua Soares. Loyola, 2009.
- MILES, J. (1995) *Deus: uma biografia*. Trad. José Rubens Siqueira. Companhia

- das Letras, 2009.
- MIRANDA, S. R. N. (org.) (2013) *O problema do mal: uma antologia de textos*. Poesis.
- MURCHO, D. (org.) *A ética da crença*. Trad. Vitor Guerreiro. Bizâncio, 2010.
- OPPY, G. (2020), "Ontological Arguments", *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/ontological-arguments/>.
- PALEY, W. (1802) "O relógio e o relojoeiro". Trad. L. H. Marques Segundo. URL = <https://epistemicufop.wordpress.com/2018/09/02/o-relogio-e-o-relojoeiro-por-william-paley/>.
- PITTARD, J. "Religious Disagreement". *The Internet Encyclopedia of Philosophy*, URL = <https://www.iep.utm.edu/rel-disa/>.
- PLANTINGA, A. (1974) *Deus, a liberdade e o mal*. Trad. Desidério Murcho. Vida Nova, 2012.
- PLANTINGA, A. (2002) "O argumento evolucionista contra o naturalismo". Trad. L. H. Marques Segundo. *Fundamento: revista de pesquisa em filosofia*, nº 13, 2016, pp. 154-158.
- PLANTINGA, A. (2011) *Ciência, Religião e Naturalismo: onde está o conflito?* Vida Nova, 2018.
- PLANTINGA, A. (2015) *Conhecimento e crença cristã*. Trad. S. R. N. Miranda. Academia Monergista, 2016.
- PORTUGAL, A. C. (2014) "Existência de Deus". In *Compêndio em Linha de Problemas de Filosofia Analítica*: http://compendioemlinha.letras.ulisboa.pt/wp-content/uploads/2015/07/portugal_2014_existência-de-deus_artigo.pdf.
- RACHELS, J. (2005) *Problemas da Filosofia*. Trad. Pedro Galvão. Gradiva, 2009.
- RITCHIE, J. (2008) *Naturalismo*. Trad. Fábio Creder. Vozes, 2012.
- ROWE, W. (2010) *Introdução à filosofia da religião*. Verbo, 2011.
- SIRE, J. (2009) *O universo ao lado: um catálogo básico sobre cosmovisão*. Trad. Marcelo Heberts. Editora Monergismo, 2018.
- SMART, J. J. C. (1989) *Nosso lugar no universo: uma questão de espaço e tempo*. Trad. J. E. Smith Caldas. Siciliano, 1991.
- SWINBURNE, R. (1996) *Deus existe?*. Trad. Agnaldo Cuoco Portugal. Academia Monergista, 2015.
- SWINBURNE, R. (2004) *A existência de Deus*. Trad. Agnaldo Cuoco Portugal.

Academia Monergista, 2015.
WARBURTON, N. (1995) *O Básico da Filosofia*. José Olympio, 2008.





Editora
UFPel

DISSERTATIO
FILOSOFIA