Editora UFPel

PROBLEMAS FILOSÓFICOS DAS CIÊNCIAS

RODRIGO REIS LASTRA CID & MATHEUS DOS REIS GOMES



DISSERTATIO FILOSOFIA

PROBLEMAS FILOSÓFICOS DAS CIÊNCIAS

SÉRIE DISSERTATIO FILOSOFIA

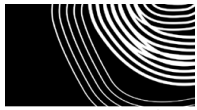
PROBLEMAS FILOSÓFICOS DAS CIÊNCIAS

Rodrigo Reis Lastra Cid
Matheus dos Reis Gomes
(Autores)

DISSERTATIO
FILOSOFIA
Pelotas, 2026.



Editora UFPel
Chefia:
Ana da Rosa Bandeira | EDITORA-CHEFE
Seção de Pré-produção:
Isabel Cochrane | ADMINISTRATIVO
Suelen Aires Böettge | ADMINISTRATIVO
Seção de Produção:
Eliana Peter Braz | PREPARAÇÃO DE ORIGINAIS
Marisa Helena Gonsalves de Moura | CATALOGAÇÃO
Anelise Heidrich | REVISÃO
Suelen Aires Böettge | ADMINISTRATIVO
Fernanda Figueredo Alves | PROJETO GRÁFICO E DIAGRAMAÇÃO
Carolina Abukawa (Bolsista) | PROJETO GRÁFICO E DIAGRAMAÇÃO
Angélica Knuth (Bolsista) | PROJETO GRÁFICO E DIAGRAMAÇÃO
Seção de Pós-produção:
Madelon Schimmelpfennig Lopes | ADMINISTRATIVO
Eliana Peter Braz | ADMINISTRATIVO





CONSELHO EDITORIAL DO NEPFIL online
Prof. Dr. João Hobuss
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 (EDITORAÇÃO)
Prof. Dr. Juliano Santos do Carmo (Editor-Chefe)

DIREÇÃO DO IFISP
Profa. Dra. Elaine Leite

CHEFE DO DEPARTAMENTO DE FILOSOFIA
Prof. Dra. Flávia Carvalho Chagas

© Série Dissertatio Filosofia, 2026.

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 organizadores. Os direitos autorais dos textos aqui publicados são de responsabilidade única e exclusiva dos organizadores do volume.

Primeira publicação em 2026 por NEPFil online e Editora da UFPel.

Dados Internacionais de Catalogação

N123 Problemas filosóficos das ciências [recurso eletrônico]. Autores: Rodrigo Reis Lastra Cid, Matheus dos Reis Gomes – Pelotas: NEPFIL Online, 2026.

478p. (Série Dissertatio Filosofia).

ISBN: 978-65-978391-9-3

1. Filosofia. 2. Ciência. I. Cid, Rodrigo Reis Lastra. II. Gomes, Matheus dos Reis.
.

COD 140







Série Dissertatio Filosofia

A Série Dissertatio Filosofia é uma iniciativa do ***Núcleo de Ensino e Pesquisa em Filosofia*** do Departamento de Filosofia da UFPel, sob o selo editorial do NEPFil online e da Editora da Universidade Federal de Pelotas. O projeto editorial tem por objetivo precípuo a publicação de livros autorais, traduções e coletâneas. O objetivo geral da série é disponibilizar materiais bibliográficos de reconhecida qualidade tanto para a divulgação da produção filosófica nacional, como para a utilização enquanto material didático e para fomentar a pesquisa filosófica.

EDITOR DA SÉRIE
Juliano Santos do Carmo (NEPFIL/UFPEL)

COMISSÃO TÉCNICA
Juliano Santos do Carmo (Editor-Chefe)

AUTORES DO VOLUME
Rodrigo Reis Lastra Cid
Matheus dos Reis Gomes

CRÉDITOS DA IMAGEM DA CAPA
Rodrigo Cid, com o uso de ferramentas de IA.

Problemas Filosóficos das Ciências

Rodrigo Reis Lastra Cid & Matheus dos Reis Gomes

Índice

Introdução 10
Filosofia da Ciência e das Ciências 18
Filosofia da Antropologia 44
Filosofia da Linguística 67
Filosofia da Matemática 101
Filosofia da Ciência da Informação 125
Filosofia da Arquivologia 172
Filosofia da Biblioteconomia 204
Filosofia da Museologia 233
Filosofia do Turismo 255
Filosofia da Engenharia e da Tecnologia 274
Filosofia da Ciência da Computação 303
Filosofia da Criptografia 319
Filosofia da Ciência Cognitiva e da Mente 341
Filosofia da Robótica e da Inteligência Artificial 371
Filosofia das Artes Digitais 393
Filosofia da Filosofia Experimental 415
Considerações Finais 435
Referências 445
Tabelas 473
Sobre os Autores 479

Introdução

Antes de começarmos propriamente a nossa jornada pelas filosofias das ciências, precisamos fazer uma digressão e nos perguntar o que é, afinal, filosofia. A pergunta parece estranha (afinal, estamos prestes a ler um livro inteiro de filosofia), mas ela é certamente necessária para compreendermos a natureza dos problemas que abordaremos aqui. Porque a filosofia não é como outras disciplinas. Se alguém se pergunta "o que é a biologia?", podemos apontar para o estudo dos seres vivos; ou "o que é a história?", podemos dizer que é o estudo do passado humano. Porém a filosofia não tem um objeto de estudo claramente delimitado como esses. Os filósofos investigam coisas tão díspares como a natureza da realidade, os limites do conhecimento, os fundamentos da moral, a estrutura do raciocínio correto, o sentido da existência, a beleza, a arte, a linguagem, a mente, a ciência, a religião, a política, o direito; a lista é longa e parece não ter fim.

O que unifica essa diversidade toda? Uma resposta tradicional é que a filosofia se caracteriza menos por seu objeto e mais por seu método, i.e., a investigação racional, sistemática e crítica das questões mais fundamentais que podemos formular. Mas isso ainda é vago. Uma outra resposta, que me parece mais útil, é a de que a filosofia lida com problemas conceituais, problemas que não podem ser resolvidos apenas coletando mais dados ou fazendo mais experimentos, mas exigem que pensemos cuidadosamente sobre os conceitos que usamos, os pressupostos que assumimos, a coerência de nossas crenças, argumentativamente.

Considere um exemplo. Um físico pode perguntar qual é a massa do elétron. Isso é uma questão empírica, pois podemos fazer experimentos para medi-la. Mas um filósofo da física pode perguntar outra coisa. "O que é massa, afinal? Massa é uma propriedade intrínseca das partículas ou uma relação com algo mais? O que significa dizer que uma partícula tem massa quando está em superposição quântica?" Essas não são questões que se resolvem com mais experimentos; elas exigem reflexão conceitual sobre o que estamos dizendo quando falamos de massa. O físico assume o conceito de massa e pergunta sobre seu valor; o filósofo pergunta sobre o próprio conceito.

Essa é uma primeira característica importante da filosofia; ela opera em um nível de reflexão de segunda ordem. Enquanto as ciências investigam o mundo, a filosofia investiga as próprias estruturas conceituais que usamos para investigar o mundo. Isso não significa que a filosofia esteja acima das ciências, ou que possa ditar verdades independentemente delas. Significa apenas que a relação é mais complexa. As ciências fornecem dados que informam a reflexão filosófica, e a reflexão filosófica clarifica os pressupostos que orientam a investigação científica.

Outra característica fundamental da filosofia é que seus problemas são persistentes. Diferentemente das ciências, onde problemas podem ser resolvidos e arquivados, os problemas filosóficos tendem a permanecer. Os gregos antigos já debatiam a natureza da justiça, o que é o conhecimento, se temos livre-arbítrio, e ainda hoje debatemos essas mesmas questões. Isso não significa que a filosofia não progride. Pelo contrário, nossos debates são hoje muito mais sofisticados, temos distinções conceituais mais refinadas, compreendemos melhor as implicações de cada posição. Mas as questões fundamentais permanecem, porque são questões sobre como conceitualizamos o mundo, e há sempre múltiplas maneiras igualmente coerentes de fazê-lo.

Os problemas filosóficos, portanto, não são como problemas de matemática, que têm respostas definitivas. Eles são mais como escolhas entre diferentes maneiras de organizar conceitualmente nossa experiência, cada uma com vantagens e desvantagens, cada uma pagando preços e colhendo benefícios. A atividade da filosofia acaba não sendo encontrar a "resposta certa" de uma vez por todas (embora a possibilidade de encontrar tal resposta seja minha esperança, ao investigar problemas filosóficos), mas mapear o espaço conceitual, clarificar as opções, avaliar argumentos, e ajudar-nos a fazer escolhas mais informadas e reflexivas.

Tradicionalmente, a filosofia ocidental se divide em algumas grandes áreas, cada uma com seu conjunto característico de problemas, a saber, a metafísica, a epistemologia, a ética e a lógica. A metafísica pergunta sobre a natureza fundamental da realidade. O que existe? Que tipos de coisas povoam o mundo? Objetos físicos, mentes, propriedades, números, Deus? Como essas coisas se relacionam entre si? Em quais categorias gerais elas se inserem? O que é causalidade? O que é espaço e tempo? Existe livre-arbítrio? Na epistemologia, nos

perguntamos sobre o conhecimento. O que é conhecimento? Como ele difere da mera crença verdadeira? Quais são as fontes do conhecimento? Podemos confiar em nossos sentidos, em nossa razão, no testemunho dos outros? O que é justificação? Temos alguma certeza? O ceticismo está correto? Além das representações da realidade e do conhecimento, a filosofia investiga também a natureza do valor ético e do valor estético. Na ética, perguntamo-nos sobre a ação correta e a vida boa. O que devemos fazer? O que significa uma ação ser moralmente correta? Existem princípios morais universais ou tudo é relativo à cultura? O que é uma vida boa e florescente? Temos obrigações para com os outros, e se sim, quais? A estética investiga a natureza da beleza, da arte e da experiência estética. O que é arte? O que torna algo belo? A experiência estética é cognitiva ou puramente emocional? Como julgar se uma obra de arte é boa esteticamente? E, finalmente, na lógica, estudamos os princípios do raciocínio correto, e ela fornece as ferramentas necessárias para avaliar argumentos de todas as áreas da filosofia, inclusive na própria lógica. O que faz um argumento ser válido? Quais são as regras da inferência dedutiva? Como distinguir raciocínios bons de ruins?

Essas são as grandes áreas tradicionais. Mas há também áreas que aplicam métodos filosóficos a domínios específicos do conhecimento ou da prática. É aqui que entram as filosofias das ciências.

A filosofia da ciência, em sentido geral, investiga questões que se aplicam a todas as ciências. O que torna uma investigação científica? Como justificamos inferências do observado para o não-observado? O que distingue ciência de pseudociência? As teorias científicas descrevem a realidade como ela é ou são apenas ferramentas úteis para fazer predições? Como a ciência muda e progride? Essas questões são transversais; aparecem, de formas diferentes, em cada campo científico, mas podem ser discutidas em um nível geral.

No entanto, cada ciência particular tem seus próprios problemas filosóficos. Ainda que, em seus próprios domínios empíricos, a física lide com partículas elementares, campos e forças, a biologia com sistemas vivos, evolução, funções, e a psicologia com a mente e o comportamento, esses diferentes objetos de estudo levantam problemas filosóficos específicos que não aparecem em outras áreas. A física quântica nos força a repensar as noções de causalidade e determinismo que a biologia pode simplesmente assumir sem problemas. A

biologia precisa lidar com noções de função e teleologia que a física ignora. A psicologia enfrenta questões sobre a relação entre mente e corpo que não têm análogo direto na química. É por isso que, ao lado da filosofia da ciência geral, temos as filosofias das ciências particulares, como a filosofia da física, filosofia da biologia, filosofia da psicologia, filosofia da economia, e assim por diante. Cada uma delas investiga os problemas conceituais que emergem no domínio específico daquela ciência, adaptando as questões gerais às peculiaridades de seu objeto e levantando questões novas que só ali fazem sentido.

Este livro caminha exatamente nesse território. Ele começa com um capítulo de filosofia da ciência em geral, estabelecendo o pano de fundo conceitual. Mas rapidamente mergulha nas especificidades de cada ciência, perguntando que problemas filosóficos particulares surgem quando estudamos as culturas? E quando investigamos as línguas e linguagens utilizadas por essas culturas? E quando as expressões linguísticas dessas linguagens se tornam informação, documento, livro, objeto museal, ou até uma experiência turística de apreciação da informação? E quando essas informações são relacionadas para criar tecnologia, máquinas, computadores? E como essa informação, dada sua natureza pode ser criada, acessada ou transferida com segurança? Os problemas são muitos e merecem nossa atenção e escrutínio criteriosos.

A ordem dos capítulos não é acidental. Há uma história sendo contada. Começamos com a antropologia, porque é ali que nos perguntamos o que é esse animal peculiar chamado humano. Não o humano biológico, que a biologia, por exemplo, já estuda. Mas o humano como ser de cultura, que cria significados, que habita mundos simbólicos, que não se contenta em apenas existir e precisa entender o que está fazendo aqui. A antropologia nos ensina que a cultura não é um adereço superficial, mas parte da constituição profunda do ser humano. Somos humanos porque somos culturais.

Mas a cultura não flutua no ar. Ela precisa de um veículo, e esse veículo é a linguagem. A linguística entra aqui como o estudo desse *medium* fundamental. A linguagem não é apenas o meio pelo qual falamos sobre o que somos; alguns suspeitam que a própria estrutura do pensamento é moldada pela língua que falamos. A filosofia da linguística pergunta o que é uma língua. O que é significado? Como símbolos se conectam ao mundo?

A linguagem, uma vez falada, se dissipa no ar. Para durar, precisa virar registro. Precisa virar informação. A ciência da informação estuda esse salto ontológico, da fala efêmera ao registro permanente. Informação é o que pode ser armazenado, recuperado, transmitido através do tempo e do espaço. É a matéria-prima de tudo o que vem depois. Mas a informação sozinha permanece abstrata. Ela precisa de corporificação. Precisa virar documento.

Arquivos, bibliotecas e museus são as instituições que dão corpo à informação, cada uma à sua maneira. Os arquivos guardam documentos como prova, i.e., registros de ações, decisões, transações que fundamentam direitos e obrigações. As bibliotecas organizam o conhecimento para ser encontrado; classificam, catalogam, disponibilizam segundo princípios que revelam visões de mundo. Os museus expõem objetos como símbolos e contam histórias com objetos, transformando artefatos em narrativas. São três modos distintos de enfrentar o mesmo problema fundamental. Como fazer o passado durar?

Mas o passado, uma vez guardado, não permanece inerte. Ele é visitado, experienciado, consumido. Por turistas. O turismo é onde essas instituições de memória encontram seu público mais amplo. O turista que entra num museu, que consulta um arquivo histórico, que percorre uma biblioteca monumental participa, muitas vezes sem perceber, do mesmo jogo hermenêutico, o jogo de dar sentido ao tempo. Mas joga de outro lugar. Não preserva, experiencia. Não cuida da memória, a consome.

Até aqui, olhamos para trás. Para o que fomos, para o que fizemos, para o que guardamos. Agora a história vira. Olhamos para frente. Para como a informação pode ser relacionada entre si e produzir os mais diversos resultados para o presente e para o futuro.

A engenharia, a ciência da computação, a criptografia, a ciência cognitiva, a robótica e a inteligência artificial são as disciplinas que “constroem o futuro”. Não o futuro como uma abstração, mas como uma realidade concreta, feita de pontes que não caem, algoritmos que decidem quem recebe crédito, moedas digitais que ninguém controla, robôs que conversam entre si, máquinas que talvez pensem. Cada uma dessas áreas traz suas próprias perguntas filosóficas urgentes. O que é um artefato tecnológico? O que é a computação? O que é o segredo criptográfico e como transmitir a informação secretamente? O que é a

mente e a natureza dos pensamentos e sua relação com a informação? O que é a inteligência? E é possível uma inteligência não biológica?

A arte digital emerge como uma síntese fascinante de muitos desses temas. Quando um artista cria uma imagem com um algoritmo, quando um NFT autentica uma obra digital, quando a criação se torna colaboração entre humano e máquina, as questões sobre autoria, autenticidade, valor e experiência estética ganham novas dimensões. A arte digital nos obriga a repensar categorias que tomávamos como garantidas. A arte digital, principalmente a que é registrada por meio de *tokens* não fungíveis (NFT) numa *blockchain*, perpassa por muitas áreas que tratamos. Ela se constitui como informação relacionada criando tecnologia, como informação guardada e registrada na memória da *blockchain*, e se apresenta como uma linguagem ou uma expressão supostamente genuinamente humana, aberta à apreciação estética, por exemplo, do turista.

E então, no final, uma última curva. A filosofia experimental chega e pergunta. Já que passamos o tempo todo examinando ciências empíricas (antropologia, linguística, ciência da informação, engenharia, computação, cognição), por que a própria filosofia não poderia se submeter a métodos empíricos? Por que não investigar, com experimentos controlados, como as pessoas realmente pensam sobre livre-arbítrio, sobre conhecimento, sobre moralidade? Não para substituir a reflexão filosófica, mas para informá-la. Para lembrar que quem filosofa é um cérebro situado, num contexto, com vieses cognitivos, e que isso, talvez, importe para as conclusões que alcançamos. De todo modo, nossa intenção aqui é mais refletir sobre a própria filosofia experimental, com seus traços empíricos, do que fazer filosofia experimental de fato.

Esta jornada não é apenas um passeio por diferentes disciplinas. Há uma tese que a orienta, a de um certo pluralismo pragmático. Não há um método único que sirva a todas as ciências, nem uma ontologia universal que capture todos os domínios de investigação. Diferentes ciências requerem diferentes ferramentas conceituais, diferentes padrões de evidência, diferentes modos de explicação. A antropologia não pode ser reduzida à física, nem a linguística à biologia, nem a ciência da informação à matemática. Cada domínio tem suas peculiaridades irredutíveis.

Mas essa diversidade não é um caos. É a resposta apropriada à complexidade do mundo que investigamos. O pluralismo que defendo não é um relati-

vismo permissivo, i.e., a ideia de que “tudo vale” ou que qualquer posição é tão boa quanto qualquer outra. Pelo contrário, reconhece que há critérios, mas que esses critérios são múltiplos, contextuais e frequentemente conflitantes. A escolha entre diferentes ontologias da cultura, diferentes métodos linguísticos, diferentes teorias da informação, diferentes políticas de acesso, diferentes designs de engenharia, todas essas escolhas podem ser avaliadas, mas a avaliação não se reduz a um único critério simples como “verdade” ou “eficiência”. Envolve, em vez disso, um julgamento ponderado que considera múltiplas virtudes, como adequação empírica, poder explanatório, consistência, simplicidade, fecundidade, e também valores éticos como justiça, autonomia, privacidade, dignidade.

O pluralismo, portanto, não é uma desistência da racionalidade, mas um reconhecimento pragmático de que a racionalidade é mais complexa do que os modelos simplistas supõem. Envolve a capacidade de navegar entre diferentes perspectivas, de integrar diferentes tipos de evidência, de pesar valores conflitantes, de exercer julgamento em situações de incerteza. É uma habilidade que se adquire com a prática reflexiva, não com a aplicação mecânica de regras.

A lista de pessoas que conversaram comigo sobre esses temas ao longo dos anos é extensa demais para caber aqui. Mas preciso mencionar, ao menos, os colegas do antigo blog de filosofia experimental, aquele projeto em Portugal que me apresentou à X-Phi, por meio do Prof. Dr. Carlos Mauro, e me fez perceber que a filosofia não precisa ser feita exclusivamente da poltrona. Preciso mencionar também os pareceristas anônimos de diversos periódicos, que com seus comentários frequentemente precisos, me obrigaram a continuamente repensar os textos aqui contidos. E preciso, evidentemente, agradecer aos estudantes. Foram eles que, com suas perguntas (aquelas perguntas aparentemente simples que revelam não termos uma resposta tão clara quanto imaginávamos), me convenceram de que este livro precisava existir.

No fundo, este livro nunca foi principalmente sobre as disciplinas acadêmicas. Foi sobre nós, sobre como nós, antropólogos, linguistas, arquivistas, bibliotecários, museólogos, profissionais do turismo, engenheiros, cientistas da computação, criptógrafos, engenheiros da robótica e desenvolvedores de IA, artistas digitais, filósofos, e sobre como cada um de nós, em nosso canto especializado, pode lidar com perguntas que transcendem os limites de nossa área.

A aposta deste livro é a de que aquela dúvida incômoda que você carrega no dia a dia de seu trabalho, aquela sensação de que algo não se encaixa perfeitamente, de que há uma pergunta mais profunda que sua disciplina não está equipada para responder sozinha, não é sinal de incompetência. É sinal de que o seu pensamento está se direcionando para o lado certo. É sinal de que você encontrou um problema genuinamente filosófico. E problemas filosóficos genuínos, embora raramente tenham soluções simples, sempre merecem ser pensados com cuidado. Esperamos que este livro ajude nessa tarefa.

1

Filosofia da Ciência e das Ciências

1. Por que uma Filosofia da Ciência?

A ciência funciona. Isso ninguém nega seriamente. Antibióticos curam infecções bacterianas. Aviões voam através de continentes. Smartphones computam operações que teriam exigido salas inteiras de equipamentos por décadas. A tecnologia baseada em conhecimento científico transformou o mundo de maneiras que pareceriam mágica genuína para nossos ancestrais. Mas por que funciona? O que torna a ciência diferente de outras formas de investigação, como a religião, o senso comum ou a tradição oral? Como os cientistas justificam suas crenças sobre o mundo?

Essas questões não são, elas mesmas, científicas; são filosóficas. Não se resolvem fazendo mais experimentos ou coletando mais dados ou calculando-se qualquer coisa, mas pensando cuidadosamente sobre a natureza, sobre os métodos e sobre os limites do conhecimento científico. A própria ciência não pode fundamentar a si mesma; por exemplo, não tem nada que um biólogo possa fazer com os métodos da biologia que lhe dê uma resposta para o que é a vida; a decisão sobre como compreender o conceito de vida e quais as implicações desse conceito para o uso na disciplina não podem ser determinadas por meio de nenhum experimento biológico. Vamos precisar da reflexão filosófica sobre a biologia, para conseguirmos chegar a uma boa conclusão sobre qual quadro conceitual os biólogos devem aceitar.

A filosofia da ciência consolidou-se como profissão acadêmica no século XX, mas as questões são ancestrais. Aristóteles já se perguntava o que constitui uma explicação científica adequada. E diversas respostas surgiram na história do pensamento humano para tentar responder essa pergunta. Francis Bacon (1620/2014), no século XVII, por exemplo, defendeu o método indutivo sistemá-

tico. Contrariamente, René Descartes (2004), contemporâneo de Bacon (1620/2014), defendeu o método dedutivo racionalista. No século seguinte, David Hume (1748/1999) demonstrou que a indução enfrenta um problema lógico profundo; e Immanuel Kant (2001) tentou salvá-la através de uma teoria transcendental do conhecimento. Embora John Stuart Mill (1843), no século XIX, tenha tentado sistematizar os métodos de investigação causal que ainda influenciavam a pesquisa empírica, foi apenas no século XX que a filosofia da ciência explodiu como campo especializado, tentando trazer questões e fundamentos para as revoluções na física da relatividade e da mecânica quântica, para a formalização rigorosa da lógica matemática, e para a profissionalização institucional da ciência como empreendimento social.

Este capítulo cumpre uma função especial no livro. Ele não examina a filosofia de uma ciência particular, mas reflete sobre o que torna todas as ciências ciências. Estabelece o cenário conceitual para todos os capítulos seguintes. As questões que examinaremos aqui, sobre indução, demarcação, realismo científico e mudança teórica, reaparecerão, transformadas, mas reconhecíveis, em cada uma das disciplinas que visitaremos. A antropologia enfrentará o problema de como generalizar legitimamente a partir de estudos de caso etnográficos. A linguística enfrentará o problema de como demarcar a linguística científica da gramática tradicional prescritivista. A ciência da informação enfrentará o problema do realismo sobre entidades informacionais abstratas. A arquivologia enfrentará o problema da mudança de paradigmas na teoria da gestão documental.

Há, portanto, uma tensão produtiva entre a unidade e a diversidade das ciências. Por um lado, todas as ciências compartilham certos problemas comuns, como o de justificar inferências do observado para o não-observado, o de demarcar uma teoria ou área como não científica, o de entender a relação entre teoria e realidade, o de explicar e avaliar a mudança de quadro teórico. Por outro lado, cada ciência possui suas peculiaridades irredutíveis. A antropologia lida com significados culturais; a física com partículas elementares; a biologia com sistemas vivos; a ciência da informação com registros e representações. A filosofia da ciência em geral ilumina esses problemas comuns; e a filosofia de cada ciência particular deve adaptar as lições gerais às suas circunstâncias específicas.

Este capítulo, portanto, não é um fim em si mesmo, mas um ponto de partida. Fornece as ferramentas conceituais que serão usadas, refinadas e às vezes contestadas nos capítulos seguintes. Também estabelece o tom intelectual do livro, que seria um forte compromisso com a clareza conceitual, uma sincera honestidade intelectual sobre as dificuldades genuínas que encontraremos, e uma total humildade sobre as nossas certezas (especialmente as aparentemente mais sólidas) e a nossa falibilidade.

A caracterização da filosofia da ciência como reflexão sobre os pressupostos das ciências particulares conecta-se com a distinção de Rudolf Carnap (1950) entre questões internas (feitas dentro de um quadro conceitual) e questões externas (sobre o próprio quadro); enquanto os cientistas fazem perguntas internas, quais são as leis do movimento, qual é a estrutura do DNA, a filosofia da ciência faz perguntas externas sobre os quadros conceituais que tornam tais perguntas possíveis. Essa distinção, ainda que contestada, delimita com certa precisão o domínio da investigação filosófica da ciência.

2. O Problema da Indução

A ciência parece depender fundamentalmente da indução. Observamos que o sol nasceu todos os dias desde que há registros históricos e astronômicos, e inferimos que nascerá amanhã. Testamos a lei da gravitação em laboratório milhares de vezes sob condições controladas, e generalizamos para todo o universo. Essa inferência do observado para o não-observado é a indução. Sem ela, muito provavelmente a ciência empírica seria impossível. Mas David Hume (1748/1999) demonstrou, no *Tratado da Natureza Humana* e depois na *Investigação sobre o Entendimento Humano*, que a indução carece de justificação lógica adequada, e tal *problema da indução* permanece ainda sem solução satisfatória três séculos depois.

Hume (1748/1999) argumentou da seguinte forma. As inferências indutivas pressupõem que o futuro será similar ao passado, i.e., que os padrões observados até agora continuarão valendo. O sol nasceu sempre no passado, logo nascerá no futuro. Mas por que deveríamos assumir a uniformidade da natureza? Não podemos provar isso logicamente. A proposição de que a natureza é uniforme não é uma verdade lógica; sua negação não implica contradição. É

concebível logicamente, sem autocontradição, que as leis naturais mudem radicalmente amanhã. Logo, a uniformidade não é necessariamente verdadeira *a priori*. Alternativamente, podemos justificá-la empiricamente, então? Não, porque isso seria circular. Usaríamos a indução para justificar a indução, pois diríamos que a natureza foi uniforme no passado, logo será uniforme no futuro. Mas essa inferência é ela mesma indutiva, pressupondo justamente o que precisa ser justificado. É um círculo vicioso manifesto.

O argumento de Hume (1748/1999) pode ser formalizado com precisão. Seja U o princípio da uniformidade da natureza. Uma justificação dedutiva de U exigiria que U fosse uma verdade lógica ou derivável de axiomas autoevidentes, o que Hume (1748/1999) nega. Uma justificação indutiva de U teria a forma "U valeu no passado, logo valerá no futuro", que pressupõe U, tornando o argumento circular. A exaustividade dessa disjunção (dedução ou indução) é ela mesma um pressuposto do argumento humeano que alguns filósofos contestam, propondo formas de inferência ampliativa que não se enquadram na dicotomia.

Portanto, a indução não tem justificação dedutiva nem indutiva. Ela parece ser um hábito psicológico profundamente enraizado, e não um raciocínio validado por princípios lógicos. Contudo, a ciência constantemente confia na indução em cada generalização empírica que faz, em cada lei natural que postula. Isso constitui um problema filosófico sério, pois, se não há justificativa para a indução, e a indução justifica as teses científicas, então pareceria que não há justificativa para as nossas teses científicas (pelo menos para as fundadas na indução). Várias tentativas de resolver esse problema foram propostas ao longo dos séculos; nenhuma é plenamente convincente.

Uma resposta natural é o probabilismo. A indução não garante a verdade com certeza absoluta, mas torna as conclusões prováveis. Quanto mais casos observamos em variadas circunstâncias, mais provável que a generalização seja verdadeira. O sol nasceu milhões de vezes em observações registradas no passado; logo é extremamente provável que nasça amanhã. A repetição supostamente aumentaria a probabilidade.

Mas Hume (1748/1999) já antecipou essa resposta. Por que as observações passadas tornam o futuro provável? Apenas se assumirmos que o passado é um guia confiável para o futuro, i.e., que a natureza é uniforme através do

tempo. Mas isso é precisamente o que precisa ser justificado. O probabilismo não resolve o problema fundamental, apenas o reformula em termos de graus de crença ao invés de certeza absoluta, e não dá razões independentes para aceitarmos sua tese central.

Carnap (1950) tentou desenvolver uma lógica indutiva rigorosa e formal. Ele propôs medidas de confirmação, quantificando precisamente o quanto uma evidência apoia uma hipótese. Mas seu sistema dependia de escolhas arbitrárias sobre distribuições de probabilidade *a priori* com relação a como distribuímos nossas crenças iniciais antes de observar qualquer evidência. Diferentes escolhas levam a diferentes graus de confirmação para a mesma evidência. Sem um princípio objetivo e não-arbitrário para escolher, a lógica indutiva de Carnap (1950) não resolve o problema humeano fundamental.

Popper (1934/1959) adotou uma estratégia radical. Ele abandonou completamente a indução. A ciência, argumentou em *A Lógica da Descoberta Científica* (1934), não procede indutivamente, mas por conjecturas ousadas e refutações decisivas. Os cientistas propõem hipóteses audaciosas que vão além dos dados conhecidos, deduzem consequências observáveis dessas hipóteses, e as testam empiricamente. Se as observações contradizem as predições, a hipótese é falsificada conclusivamente. Se as observações concordam com as predições, a hipótese sobrevive, mas não é confirmada ou verificada; apenas ainda não foi refutada. A ciência nunca justifica teorias positivamente, apenas elimina erros através de falsificações. Isso evita o problema da indução, porque nos diz que a ciência não procede por indução, mas por falsificação de hipóteses.

Mas o falsificacionismo enfrenta problemas próprios, tanto lógicos quanto históricos. Um deles é que as observações não falsificam as teorias isoladamente. As teorias são testadas conjuntamente com múltiplas hipóteses auxiliares sobre os aparatos experimentais, as condições iniciais, as teorias de fundo. Se uma predição falha, o que está errado? A teoria principal, ou uma hipótese auxiliar, ou o instrumento de medição? As observações parecem já carregadas de teorização. Isso é a tese de Duhem-Quine; as teorias enfrentam o tribunal da experiência holisticamente, como uma rede interconectada, não isoladamente. Logo, a falsificação não é tão simples e conclusiva quanto Popper (1934/1959) sugeria.

Além disso, os cientistas frequentemente mantêm teorias apesar de falsificações ou anomalias persistentes. Newton sabia que a órbita de Mercúrio apresentava anomalias em relação às predições da gravitação newtoniana, mas não abandonou sua teoria. E, de fato, ele estava correto em manter a confiança em sua teoria, pois a anomalia foi eventualmente explicada pela relatividade geral, que mostrou não um erro fundamental de Newton, mas algumas limitações de escopo. Se os cientistas abandonassem as teorias ao primeiro contra-exemplo, a ciência não progrediria. É preciso haver espaço metodológico para defender teorias produtivas contra anomalias.

E, mais fundamentalmente, mesmo o falsificacionismo depende implicitamente da indução. Quando afirmamos que uma observação falsifica uma teoria, assumimos que a observação é confiável, que os instrumentos funcionam adequadamente, que as condições experimentais são apropriadas. Todas essas assunções de fundo dependem de generalizações indutivas de casos passados. Logo, o falsificacionismo não escapa completamente da indução, apenas a desloca para níveis menos visíveis da prática científica.

Goodman (1955) apresentou um problema adicional que complica ainda mais o panorama, a saber, *o novo enigma da indução*. Suponha que observamos esmeraldas até agora e todas foram verdes. Induzimos naturalmente que todas as esmeraldas são verdes. Mas defina uma nova propriedade, “verdul”, que seria verde se observado antes do tempo t (digamos, amanhã), azul caso contrário.

A evidência disponível, a saber, todas as esmeraldas observadas até agora, apoia igualmente as generalizações “todas as esmeraldas são verdes” e “todas as esmeraldas são verduis”. Contudo, essas generalizações fazem predições radicalmente opostas sobre esmeraldas observadas após t. A primeira prediz esmeraldas verdes, a segunda azuis.

O enigma de Goodman (1955) tem implicações diretas para a aprendizagem de máquinas. Qualquer conjunto finito de dados de treinamento é igualmente compatível com infinitas hipóteses, incluindo hipóteses “perversas” (*grue-like*) que fazem predições radicalmente diferentes para dados futuros. O problema do viés de indução (*inductive bias*) em aprendizagem computacional é precisamente a versão técnica do enigma de Goodman (1955); algum viés a

priori sobre quais hipóteses preferir é inevitável, e a justificação desse viés recai no mesmo problema filosófico que Goodman (1955) identificou.

A indução não determina qual generalização aceitar sem assumir que certas propriedades (como "verde") são projetáveis (legitimamente generalizáveis), enquanto outras (como "verdul") não são. Mas o que torna uma propriedade projetável? Parece que é simplesmente o fato de termos usado essa propriedade em generalizações bem-sucedidas no passado. Mas isso é circular novamente, pois usamos a indução passada para justificar a indução futura.

Salmon (1967) propôs uma justificação pragmática sofisticada. A indução não pode ser justificada logicamente *a priori*, mas se há regularidades na natureza, a indução as descobrirá eventualmente através de ajustamento e autocorreção. Se não há regularidades, nenhum método possível funcionará. Logo, racionalmente devemos usar a indução, pois é nossa melhor aposta. É claro que o problema aqui é que isso pressupõe que há regularidades, justamente o que precisa ser justificado. Mas seu ponto aqui é ser uma versão sofisticada da aposta de Pascal, e não uma solução do problema lógico fundamental.

Naturalistas como Quine (1969) argumentam que o problema da indução é um pseudoproblema gerado por exigências fundacionistas irrealistas. A indução seria parte constitutiva de nossa natureza cognitiva, moldada por milhões de anos de evolução biológica. Ela funciona porque organismos com tendências indutivas mal calibradas não sobreviveram para reproduzir. Não precisamos de uma justificação *a priori* transcendente; a justificação é empírica e evolucionária. A indução justifica-se por seus resultados ao longo da história natural.

Mas essa resposta não satisfaz completamente o cético filosófico. A evolução nos dotou de hábitos cognitivos adaptativamente úteis, mas "útil para sobrevivência" não é idêntico a "epistemicamente justificado" ou "confiável para alcançar a verdade". A indução pode ser evolutivamente adaptativa sem ser epistemicamente justificada. O problema de Hume (1748/1999) persiste, i.e., como justificamos racionalmente a transição do observado para o não-observado?

A lacuna entre adaptabilidade evolutiva e confiabilidade epistêmica é o núcleo do problema da epistemologia evolucionária. Plantinga (1993) formulou o argumento da deferrência naturalista (*evolutionary debunking argument*) na direção oposta, dado que se nossas faculdades cognitivas foram selecionadas

por adaptação e não por produção de verdade, temos razões para desconfiar da confiabilidade dessas faculdades, inclusive da própria crença no naturalismo evolucionário. A resposta de Quine (1969) pressupõe que adaptação e verdade covariam suficientemente, o que é exatamente o ponto em questão.

Minha posição, compartilhada por um crescente número de epistemólogos contemporâneos, é que o problema da indução não admite solução satisfatória nos termos fundacionistas tradicionais. Não podemos justificar a indução dedutivamente porque ela não é uma dedução. É inferência ampliativa que vai além da informação contida nas premissas. Não podemos justificá-la indutivamente sem circularidade viciosa. Não podemos justificá-la empiricamente sem pressupor o que queremos justificar.

Talvez precisemos aceitar a indução como uma capacidade epistêmica básica, não fundamentada em algo mais básico, mas também não carecendo de tal fundamentação para ser legítima. Isso é desconfortável para fundacionistas (como eu) que buscam justificações últimas, mas talvez reflita simplesmente como a cognição humana funciona, havendo pontos de partida não-justificados que tornam a justificação possível.

Isso não torna a ciência irracional ou arbitrária. Torna a racionalidade mais modesta e menos triunfalista. A racionalidade epistêmica não é apenas dedução rigorosa a partir de axiomas auto-evidentes. Ela inclui confiar em intuições bem calibradas, fazer inferências não-demonstrativas mas frutíferas, revisar crenças à luz da experiência acumulada, balancear virtudes teóricas como simplicidade, poder explanatório e coerência interna e externa. A indução seria parte integrante desse pacote cognitivo complexo, não algo que precisaria ser reduzido à dedução para ser epistemicamente legítimo.

A noção de "virtudes teóricas" (*theoretical virtues*) que aqui aparece, simplicidade, poder explanatório, coerência, é tematizada sistematicamente por Thomas Kuhn (1977) em "Objectivity, Value Judgment, and Theory Choice", onde as apresenta como critérios que orientam a escolha entre teorias sem determiná-la algoritmicamente. A indução funcionaria como parte desse conjunto de virtudes epistêmicas, cujo exercício requer julgamento e não computa-se mecanicamente.

O problema da indução reaparecerá de forma particularmente urgente quando examinarmos o aprendizado de máquinas, nos capítulos sobre ciência

da computação e sobre robótica e inteligência artificial. Os algoritmos de aprendizado supervisionado são, essencialmente, procedimentos indutivos automatizados, pois generalizam a partir de dados de treinamento observados, para fazer predições sobre dados não-observados. O problema de Hume (1748/1999) manifesta-se concretamente na questão de por que um modelo treinado em amostras passadas deveria funcionar bem em amostras futuras. A teoria do aprendizado computacional desenvolve ferramentas matemáticas para quantificar essa confiabilidade, mas as questões filosóficas fundamentais sobre a legitimidade da indução permanecem.

3. O Problema da Demarcação

O que torna uma disciplina uma ciência, em vez de pseudociência ou mito? Intuitivamente, parece que deveria haver uma resposta clara e precisa. A física é ciência, a astrologia não é. A biologia evolutiva é ciência, o criacionismo da Terra jovem não é. A medicina baseada em evidências é ciência, a homeopatia não é. Mas quando tentamos articular um critério preciso, geral e não-arbitrário que separe a ciência da não-ciência, algumas dificuldades profundas emergem. Este é o *problema da demarcação*, e sua história no século XX se constitui de uma série de propostas filosóficas ambiciosas seguidas de contra-exemplos devastadores.

Os positivistas lógicos do Círculo de Viena, como Carnap (1928) e Neurath, Carnap & Morris (1938), propuseram o verificacionismo como critério de demarcação. Os enunciados científicos são verificáveis empiricamente; enunciados metafísicos ou não científicos não são. Se uma proposição não pode, pelo menos a princípio, ser verificada ou falsificada pela observação empírica, ela carece de significado cognitivo; é literalmente sem sentido, mero ruído verbal. Isso excluiria a metafísica tradicional, a teologia especulativa e até algumas formas de éticas normativas, deixando apenas a ciência empírica e as verdades lógico-matemáticas como portadoras genuínas de conteúdo cognitivo.

Mas o verificacionismo é auto-refutante de maneira espetacular! O próprio princípio de verificação de *apenas proposições verificáveis empiricamente terem significado cognitivo* não é ele mesmo verificável empiricamente, nem é uma verdade lógica analítica. Logo, pelos seus próprios critérios rigorosos, o princí-

pio de verificação é sem sentido. É uma contradição performativa. Chega a ser engraçado o fato de eles não terem percebido isso.

Além disso, o verificacionismo é excessivamente restritivo mesmo para as ciências que pretendia fundamentar. As leis científicas universais não são verificáveis conclusivamente. A lei de que todos os corvos são pretos não pode ser verificada, pois não podemos observar todos os corvos passados, presentes e futuros. Leis causais, disposições e contrafactuais resistem à verificação direta. O verificacionismo, se aplicado rigorosamente, excluiria muito do que intuitivamente reconhecemos como ciência legítima. Dessa forma, tal teoria não poderia estar correta.

Popper (1934/1959) propôs o falsificacionismo como critério superior. As teorias científicas são falsificáveis em princípio; as teorias pseudocientíficas não são. A astrologia faz predições tão vagas e ajustáveis que nunca arriscam falsificação genuína. A psicanálise freudiana, como um outro exemplo, tal como argumentou Popper (1963), poderia acomodar qualquer comportamento observado através de reinterpretações *ad hoc*. O mesmo para as formas vulgares de marxismo, que predizem uma revolução proletária inevitável, mas quando ela não acontece conforme predito, ajustam a teoria retrospectivamente para explicar o atraso. Essas teorias são imunes à refutação empírica, logo não são genuinamente científicas, para Popper (1963).

Por outro lado, num contraste dramático, Einstein fez uma predição arriscada e precisa de que a luz de estrelas distantes seria curvada pelo campo gravitacional do Sol de uma quantidade específica calculável. A expedição de Eddington em 1919 para observar um eclipse solar poderia ter refutado conclusivamente a relatividade geral se não observasse o encurvamento predito na posição aparente das estrelas. A expedição observou o desvio predito, e a teoria sobreviveu ao teste arriscado. Mas o ponto crucial de Popper (1934/1959) é que a teoria estava genuinamente em risco, pois havia condições observacionais especificáveis de antemão que teriam falsificado a teoria, se tivessem ocorrido. Isso é a marca distintiva da ciência genuína para ele.

O caso da expedição de Eddington tem sido revisitado criticamente por historiadores da ciência; análises posteriores dos dados originais sugerem que havia considerável margem de erro e que a confirmação foi menos inequívoca do que a narrativa canônica indica. O caso ilustra precisamente a tese de Du-

hem-Quine de que a interpretação dos dados dependia de hipóteses auxiliares sobre os instrumentos e as condições meteorológicas, e a decisão de tratar os resultados como confirmatórios envolveu julgamentos que não foram puramente lógicos.

O falsificacionismo possui um apelo intuitivo considerável. Ele captura algo importante sobre a atitude científica, que seria a disposição para submeter as teorias a testes rigorosos, onde o fracasso é possível, por meio de conjecturas audaciosas, feitas com a coragem de arriscar serem falsas, a fim de tentar descobrir a verdade. Porém, como critério de demarcação preciso e universal, o falsificacionismo enfrenta problemas sérios. Um deles, como vimos na seção anterior, é que as teorias não são falsificáveis isoladamente. Sempre testamos sistemas teóricos completos incluindo hipóteses auxiliares sobre a funcionalidade dos nossos instrumentos, sobre quais são as condições de fundo, sobre as teorias conectadas com aquela que estamos testando. Logo, qualquer teoria pode ser protegida contra falsificação ajustando as hipóteses auxiliares; portanto a distinção entre ciência e pseudociência não pode residir na falsificabilidade em princípio, mas antes na atitude concreta dos praticantes, na sua disposição para aceitar falsificações genuínas em vez de uma propensão para ajustes *ad hoc*.

Outro ponto é que a falsificação é difícil quando testamos uma teoria probabilística, porque o resultado não sair o mais provável ou sair o menos provável não falsifica a teoria. Como muitas teorias científicas maduras fazem predições probabilísticas, não determinísticas, cabe falarmos sobre esse ponto. A mecânica quântica, por exemplo, prediz as probabilidades dos resultados das medições, e não resultados específicos determinados. Se um resultado altamente improvável ocorre, isso não falsifica a teoria, pois era apenas improvável, não impossível. Onde traçar a linha entre o improvável e a falsificação? Não há critério objetivo e não-arbitrário. A falsificação probabilística é uma questão de julgamento, não de lógica pura.

Uma outra falha do falsificacionismo é que algumas teorias genuinamente pseudocientíficas fazem predições testáveis que são repetidamente falsificadas, mas seus proponentes as ignoram sistematicamente. O criacionismo da Terra jovem prediz que a Terra tem menos de 10.000 anos. Múltiplas linhas independentes de evidência (datação radiométrica, estratigrafia geológica, dendrocrono-

logia, camadas de gelo polar) convergem em bilhões de anos. Mas os criacionistas mantêm a teoria, ajustando as hipóteses auxiliares *ad infinitum* (Deus criou a aparência de idade, as taxas de decaimento mudaram etc). O problema não é a falta de falsificabilidade em princípio, mas a recusa de aceitar a falsificação na prática.

Lakatos (1970, 1978) sofisticou Popper (1934/1959) com a metodologia dos programas de pesquisa científica. Um programa de pesquisa possui um núcleo duro de assunções centrais, protegido contra falsificação por decisão metodológica, e um cinturão protetor de hipóteses auxiliares ajustáveis. O programa é progressivo se prediz fatos novos surpreendentes que são subsequentemente confirmados; é degenerativo se apenas acomoda fatos já conhecidos através de ajustes retrospectivos. A ciência genuína emprega programas progressivos; a pseudociência opera com programas permanentemente degenerativos.

Isso é metodologicamente mais sofisticado que o falsificacionismo ingênuo. Permite que teorias sejam racionalmente protegidas contra anomalias temporárias, algo que corresponde muito melhor à prática científica histórica real. Newton manteve sua teoria gravitacional apesar da anomalia da órbita de Mercúrio, e estava correto em fazê-lo. Mas Lakatos (1970) não fornece um critério de demarcação claro e operacional. Quando exatamente um programa transita de temporariamente anômalo para permanentemente degenerativo? Não há resposta objetiva e temporal precisa. O julgamento contextual é necessário, e julgamentos razoáveis podem divergir.

Kuhn (1962) argumentou que a distinção entre ciência e pseudociência não é primariamente metodológica, mas sociológica e histórica. A ciência madura possui um paradigma aceito consensualmente pela comunidade científica relevante. O paradigma define os problemas legítimos, os métodos apropriados e os padrões aceitáveis de solução. Os cientistas trabalham dentro de tal paradigma fazendo o que Kuhn (1962) chama de ciência normal, isto é, resolvendo quebra-cabeças, refinando medições, estendendo aplicações. As revoluções científicas ocorrem quando os paradigmas começam a apresentar anomalias, ocasionalmente substituindo paradigmas exaustos por novos. A pseudociência, por contraste, não possui um paradigma consensual, além de carecer de comunidade madura e estável e não resolver os quebra-cabeças e dificuldades acumulados dentro da sua própria tradição.

Essa análise sociológica captura algo importante negligenciado por abordagens puramente lógicas, a de que a ciência é fundamentalmente um empreendimento social e comunitário, não apenas um método individual de investigação. Entretanto, essa ideia torna a demarcação relativa, histórica e contingente de maneiras realmente desconfortáveis. A alquimia medieval possuía uma comunidade de praticantes e métodos compartilhados. Isso a tornava ciência naquela época? Tal como a astrologia helenística, que tinha tradição estabelecida. A linha entre ciência e pseudociência torna-se excessivamente turva, aceitando no domínio de ciência mais coisas do que gostaríamos de aceitar.

A tensão entre o reconhecimento da dimensão social da ciência e a manutenção de critérios normativos de demarcação é o núcleo do debate entre o programa forte da sociologia do conhecimento científico e os seus críticos. O programa forte sustenta que o conteúdo das crenças científicas deve ser explicado causalmente pelos mesmos tipos de fatores sociais, independentemente de serem verdadeiras ou falsas. A resposta filosófica padrão é que essa simetria explicativa compromete a possibilidade de distinguir entre crença bem justificada e mero consenso social, distinção que qualquer epistemologia normativa da ciência precisa preservar.

Feyerabend (1975) radicalizou ainda mais e argumentou que não existe um método científico universal. Cientistas historicamente bem-sucedidos usaram qualquer estratégia que funcionasse, de propaganda retórica até falsificação de dados, violação de métodos aceitos, intuição mística. Galileu usou uma retórica persuasiva e telescópios não-calibrados para converter seus adversários. Isso foi essencial para a revolução científica. Tentar codificar o método científico em regras metodológicas rígidas restringiria a criatividade e bloquearia a inovação. O único princípio metodológico defensável, concluiu provocativamente Feyerabend (1975), é "*anything goes*", i.e., tudo vale, não há regras universais.

Feyerabend (1975) é frequentemente caricaturado e mal interpretado. Ele não estava defendendo um irracionalismo niilista, mas atacando o imperialismo metodológico, i.e., a imposição arrogante de um único conjunto de métodos sobre todas as formas de investigação. Diferentes campos com diferentes objetos precisam de diferentes métodos. Não há receita universal. Mas levada ao extremo, essa posição dissolve completamente a demarcação. Se metodologicamente tudo vale, então a astrologia e a parapsicologia são tão legítimas quan-

to a física. Poucos, incluindo o próprio Feyerabend (1975) em momentos mais sóbrios, aceitam essa conclusão radical.

Laudan (1983) concluiu que o problema da demarcação é um pseudoproblema filosófico. Não há, nem nunca haverá, um critério único, preciso e não-arbitrário que separe universalmente a ciência da não-ciência. É melhor abandonar a busca e focar pragmaticamente em avaliar alegações específicas por seus méritos epistêmicos concretos, independentemente do rótulo disciplinar. Algumas alegações apresentadas como científicas são epistemicamente ruins; algumas alegações rotuladas como pseudocientíficas possuem mérito evidencial. Os rótulos não importam fundamentalmente; apenas a qualidade da evidência, do argumento e da justificação.

Laudan (1983) pode estar correto que o problema filosófico tradicional de buscar condições necessárias e suficientes para "ciência" não tem solução. Mas o problema prático persiste urgentemente. Educadores precisam decidir o que ensinar em aulas de ciência. Tribunais precisam decidir no que acreditar como provas científicas. Agências de financiamento precisam decidir que propostas de pesquisa merecem recursos públicos. Jornalistas precisam decidir que alegações tratar como cientificamente legítimas. Cidadãos precisam decidir em que experts confiar. Sem algum critério de demarcação, mesmo imperfeito, essas decisões parecem arbitrárias ou meramente políticas.

Minha posição, compartilhada por muitos filósofos contemporâneos, é que "ciência" é um conceito de família ou *cluster concept* no sentido de Wittgenstein (1953/2009). Não há condições individualmente necessárias e conjuntamente suficientes que definem a ciência, mas uma família de características típicas que frequentemente andam juntas, como a testabilidade empírica sistemática, o uso de controles experimentais e a metodologia rigorosa, a quantificação e a medição precisa, a busca por leis gerais ou padrões explicativos, a revisibilidade das teorias à luz de evidências novas, a comunicação pública e o escrutínio crítico, uma comunidade de praticantes com padrões compartilhados, o progresso cumulativo em poder preditivo e explicativo, a integração com outras áreas de conhecimento estabelecido. Quanto mais dessas características uma disciplina exibe, mais claramente científica ela é. Física e biologia molecular exibem quase todas. Astrologia e psicanálise freudiana exibem muito poucas. Alguns casos são limítrofes, como a economia neoclássica, certas áreas da

psicologia social, algumas subdisciplinas da antropologia. Isso não é uma falha conceitual, mas reflete a realidade dos conceitos do mundo real, que possuem fronteiras graduais, e não linhas precisas.

O problema da demarcação reaparecerá em cada ciência particular. A antropologia debaterá se é ciência social ou humanidade interpretativa. A linguística chomskyiana reivindicou status de ciência natural, desafiando tradições descritivas. A ciência da informação luta para se estabelecer como disciplina científica distinta da tecnologia da informação. O problema não desaparece, apenas se manifesta diferentemente em cada domínio.

4. O Problema do Realismo Científico

Consigamos ou não desvendar os problemas filosóficos da justificação da ciência, uma pergunta muito importante fica no ar sobre a natureza das teorias científicas. Elas descrevem a realidade como ela realmente é, independentemente de nossas mentes e observações, ou são apenas ferramentas instrumentais úteis para organizar experiências e fazer predições? Os elétrons, genes e campos gravitacionais existem realmente "lá fora", ou são ficções convenientes que sistematizam dados empíricos? Esta é a questão do realismo científico, sobre se devemos pensar as entidades postuladas pelas ciências como reais, um dos debates mais persistentes e profundos da filosofia da ciência.

O realismo científico afirma que teorias científicas maduras e bem-confirmadas são aproximadamente verdadeiras em suas afirmações sobre o mundo, e as entidades teóricas que postulam (elétrons, fótons, DNA, tectônica de placas) existem realmente. Quando a física quântica afirma que elétrons possuem carga elétrica negativa, massa de repouso específica e spin intrínseco, está descrevendo características reais de partículas reais, não meramente sistematizando dados experimentais.

O argumento central para o realismo é o argumento do milagre, articulado influentemente por Putnam (1975), de que o sucesso espetacular da ciência, tanto preditivo quanto tecnológico, seria um milagre inexplicável, se as teorias científicas fossem radicalmente falsas sobre a estrutura do mundo. O GPS funciona com precisão de metros porque a relatividade geral está correta sobre dilatação temporal gravitacional. Os eletrônicos semicondutores funcionam

porque a mecânica quântica está correta sobre o comportamento de elétrons em cristais. Os antibióticos funcionam porque a bioquímica está correta sobre estrutura molecular das proteínas bacterianas. Se essas teorias fossem completamente falsas em suas afirmações sobre entidades inobserváveis, o sucesso tecnológico derivado delas seria uma coincidência cósmica milagrosa. Logo, a melhor explicação, a única explicação não-miraculosa, do sucesso científico é que as teorias são aproximadamente verdadeiras.

Mas anti-realistas possuem contra-argumentos poderosos. Laudan (1981) apresentou a devastadora meta-indução pessimista, segundo a qual a história da ciência está repleta de teorias extraordinariamente bem-sucedidas que eram, contudo, radicalmente falsas sobre suas entidades teóricas postuladas. A teoria do flogisto explicava combustão e calcinação sistematicamente, mas o flogisto não existe. A teoria do éter luminífero explicava propagação de luz e fenômenos ópticos, mas o éter não existe. A teoria do calórico explicava transferência de calor e termodinâmica, mas o calórico não existe. A teoria atômica de Dalton explicava proporções químicas, mas seus átomos indivisíveis não existem. Se as teorias bem-sucedidas do passado eram sistematicamente falsas sobre sua ontologia, por que acreditar que as teorias atuais bem-sucedidas são verdadeiras? A indução sobre a história da ciência sugere pessimisticamente que as teorias atuais também serão eventualmente substituídas, suas entidades reveladas como ficções úteis mas irreais. O realismo parece ingênuo à luz da história.

Os realistas responderam sofisticando a sua posição. O realismo estrutural distingue o núcleo estrutural de teorias (relações matemáticas, equações fundamentais) dos detalhes ontológicos específicos sobre que entidades implementam essa estrutura. O flogisto estava errado como substância, mas algo sobre a combustão estava correto, a saber, a estrutura das reações de oxidação. O éter estava errado como meio material, mas algo sobre a propagação das ondas eletromagnéticas estava correto, a saber, as equações de Maxwell. As teorias antigas capturaram a estrutura matemática importante, apesar de terem uma ontologia equivocada. Teorias atuais também podem ter uma ontologia parcialmente errada, mas capturam estruturas relacionais verdadeiras. O realismo estrutural evita a meta-indução pessimista focando na estrutura matemática

conservada através das mudanças teóricas, não nas entidades específicas postuladas.

O realismo estrutural apresenta-se tanto na versão epistêmica (Worrall, 1989), que sustenta apenas que podemos conhecer a estrutura das teorias, permanecendo agnósticos quanto à natureza das entidades, quanto na versão ôntica (Ladyman; French 2003), que vai além e sustenta que a estrutura é o que existe, sendo as entidades individuais derivadas ou eliminadas. A versão ôntica gera dificuldades filosóficas próprias, em especial quanto à individuação dos objetos físicos e à compatibilidade com a física quântica, onde objetos indistinguíveis não são indivíduos no sentido clássico.

Bas van Fraassen (1980) propôs o empirismo construtivo como alternativa sofisticada ao realismo. O objetivo da ciência não é a verdade sobre entidades inobserváveis, mas adequação empírica, i.e., "salvar os fenômenos" observáveis. As teorias são aceitas quando são empiricamente adequadas (fazem predições corretas sobre observáveis), não necessariamente quando são verdadeiras sobre inobserváveis. Podemos permanecer agnósticos sobre se os elétrons realmente existem, enquanto confiamos que as teorias sobre os elétrons fazem predições empiricamente corretas.

Essa posição evita o compromisso ontológico com inobserváveis, enquanto permite usar as teorias científicas normalmente para predição e aplicação tecnológica. Mas enfrenta desafios próprios. Não explica satisfatoriamente por que as teorias são empiricamente adequadas, se não são aproximadamente verdadeiras. O sucesso preditivo sistemático parece requerer explicação, e a explicação mais natural é a verdade aproximada. E a distinção observável/inobservável que van Fraassen (1980) assume é problemática. O que conta como observável? Van Fraassen (1980) diz que é o que pode ser detectado sem instrumentos, apenas com sentidos humanos não-aumentados. Mas essa linha é arbitrária e antropocêntrica. As luas de Júpiter são observáveis através de telescópio, bactérias através de microscópio, elétrons através de câmara de bolhas. Por que desenhar a linha ontológica entre observação a olho nu e observação instrumentalmente mediada? Os instrumentos científicos estendem as capacidades perceptuais humanas, não criam um reino metafísico separado.

Minha posição é um realismo modesto e seletivo. Acredito que teorias científicas maduras capturam algo genuinamente real sobre a estrutura do mundo,

mas permaneço agnóstico sobre exatamente o quê. Os elétrons existem em algum sentido robusto (não são meras ficções instrumentais), mas sua natureza precisa pode ser diferente do que teorias atuais descrevem. O realismo total sobre a ontologia atual seria um otimismo excessivo e historicamente ingênuo, dada a meta-indução pessimista. Mas o anti-realismo total de tratar todas as entidades teóricas como meras construções instrumentais não explica adequadamente o sucesso da ciência e parece inconsistente com a prática científica real.

O debate realismo/anti-realismo manifestar-se-á diferentemente em cada ciência. A matemática enfrentará questões sobre o realismo platonista de números. A ciência da informação debaterá se informação é uma entidade real ou uma construção teórica. A ciência cognitiva enfrentará questões sobre o realismo de representações mentais. O framework geral orienta, mas cada domínio possui nuances próprias.

5. O Problema da Mudança Científica

A ciência muda. Teorias são substituídas, conceitos fundamentais são revisados, paradigmas inteiros colapsam. A mecânica newtoniana deu lugar à relatividade einsteiniana. A química do flogisto foi substituída pela química da oxidação. A genética mendeliana foi integrada e transformada pela biologia molecular. Como devemos entender essas mudanças? São acumulações progressivas de conhecimento ou revoluções disruptivas que rompem com o passado? Este é o problema da mudança científica, e nossas respostas moldam profundamente como compreendemos a própria natureza da ciência.

A visão tradicional, ainda predominante em livros didáticos e na imagem pública da ciência, é o cumulativismo. A ciência acumula o conhecimento progressivamente através dos séculos. Cada geração adiciona tijolo sobre tijolo ao edifício do conhecimento científico. Erros são corrigidos, lacunas são preenchidas, teorias são refinadas, mas o fundamento permanece sólido. Newton não destruiu Galileu; expandiu e sistematizou seus insights sobre o movimento. Einstein não destruiu Newton; mostrou que a mecânica newtoniana é uma aproximação válida para velocidades baixas e campos gravitacionais fracos. A ciência é um processo linear de aproximação crescente à verdade sobre o mundo.

Essa visão possui um apelo intuitivo considerável. Ela explica o sucesso tecnológico crescente, justifica a confiança pública na ciência, e apresenta uma imagem reconfortante de progresso intelectual cumulativo. Mas historiadores e filósofos da ciência mostraram que essa imagem é seriamente simplificada, se não fundamentalmente enganosa.

Kuhn (1962) desafiou radicalmente a visão cumulativista. A mudança científica, argumentou, não é primariamente cumulativa mas revolucionária. A ciência alterna entre períodos de ciência normal, onde um paradigma é consensualmente aceito, e revoluções científicas, onde o paradigma exaurido é substituído por um radicalmente novo. Os paradigmas rivais são incomensuráveis, a saber, não há um padrão neutro e independente para escolher racionalmente entre eles. A mudança é mais sociológica e psicológica do que puramente lógica ou racional.

Durante a ciência normal, os cientistas trabalham dentro de um paradigma compartilhado, que define quais problemas são significativos, quais soluções são aceitáveis, quais métodos são apropriados, que virtudes teóricas importam. Anomalias, i.e., fenômenos que não se encaixam nas predições paradigmáticas são esperadas e toleradas. Cientistas confiam que serão eventualmente resolvidas dentro do paradigma através de trabalho normal; por isso a ciência normal é conservadora por natureza; não busca novidades disruptivas mas um refinamento cumulativo dentro do paradigma.

Mas as anomalias eventualmente se acumulam. Quando persistem, instala-se uma crise de confiança no paradigma. Os cientistas, especialmente os mais jovens e menos socializados no paradigma antigo, propõem paradigmas alternativos radicais. A revolução científica ocorre quando a comunidade científica abandona coletivamente o paradigma velho e adota um novo, como nos casos de Copérnico x Ptolomeu, Newton x Aristóteles, Einstein x Newton, Darwin x Criacionismo, mecânica quântica x mecânica clássica.

Paradigmas rivais, para Kuhn (1962), são incomensuráveis, i.e., usam conceitos fundamentais diferentes, medem quantidades diferentes, valorizam virtudes teóricas diferentes, veem problemas diferentes como importantes. A massa newtoniana da quantidade constante de matéria não é idêntica à massa relativística que é dependente da velocidade. O tempo newtoniano, absoluto e universal, não é o tempo relativístico, relativo ao observador. Os cientistas em paradigmas diferentes, afirma Kuhn (1962) provocativamente, “vivem em mun-

dos diferentes". Não há uma base neutra e compartilhada para comparar paradigmas racionalmente, de forma conclusiva. A escolha entre paradigmas é, em aspectos importantes, mais uma conversão gestáltica do que uma demonstração lógica.

Essa visão relativista chocou muitos filósofos e cientistas. Se diferentes paradigmas são genuinamente incomensuráveis, como a ciência pode ser progressiva em qualquer sentido objetivo? Se a mudança de paradigma é irracional e, assim, não baseada em razões que compelem universalmente, como a ciência é epistemicamente superior a um dogma religioso ou a um modismo cultural? Kuhn (1962) foi acusado de reduzir a ciência à "psicologia de massas" ou "sociologia do conhecimento", minando qualquer alegação de objetividade.

Kuhn (1977) respondeu que a mudança não é arbitrária ou irracional, apenas não compele logicamente no sentido dedutivo rigoroso. Os cientistas valorizam critérios objetivos, como precisão preditiva, escopo explicativo, simplicidade teórica, fertilidade heurística, consistência com outras áreas estabelecidas. Novos paradigmas geralmente são superiores aos antigos nesses critérios. Mas os critérios são aplicados diferentemente por pessoas diferentes, ponderados diferentemente, interpretados contextualmente. Há razões para escolhas entre paradigmas, mas essas razões não são determinantes da escolha racional da mesma forma que as demonstrações matemáticas determinam suas conclusões.

A resposta de Kuhn (1977) à acusação de irracionalismo articula-se, em *Objectivity, Value Judgment, and Theory Choice*, com uma distinção entre critérios compartilhados e algoritmos de decisão; os cientistas compartilham os valores que devem orientar a escolha teórica (precisão, simplicidade, fecundidade, coerência, escopo), mas a aplicação desses valores é inescapavelmente interpretativa. Isso não é relativismo, mas reconhecimento de que a racionalidade científica é uma prática normativa situada, não um procedimento mecânico.

Popper (1963) criticou Kuhn (1962) ferozmente. A "ciência normal" kuhniana é dogmática, acrítica, fundamentalmente oposta ao espírito científico genuíno. A ciência verdadeira sempre questiona teorias estabelecidas, sempre busca refutações, nunca aceita dogmas inquestionáveis. Os paradigmas kuhnianos são camisas de força intelectuais que inibem a crítica saudável e protegem as teorias de testes rigorosos. A incomensurabilidade é um mito filosófico; teorias são

sempre comparáveis racionalmente através de testes empíricos decisivos, pensava Popper (1963).

Mas Kuhn (1962) tinha uma resposta para isso, a de que a ciência popperiana, na qual todo cientista constantemente tenta refutar todas as teorias, seria caótica e improdutiva. A ciência precisa de períodos de trabalho consolidado dentro de quadros conceituais estáveis, para desenvolver teorias completamente, explorar suas consequências, resolver quebra-cabeças detalhados. Revoluções constantes impediriam o progresso cumulativo dentro das tradições. A tensão dialética entre o conservadorismo paradigmático da ciência normal e a inovação disruptiva de uma revolução é produtiva, não problemática.

Lakatos (1970) tentou justamente sintetizar Popper e Kuhn (1962), através da metodologia dos programas de pesquisa científica. Os programas de pesquisa, tal como os paradigmas kuhnianos, possuem um núcleo duro protegido e orientam a pesquisa normal. Mas, diferentemente de Kuhn (1962), para Lakatos (1970), a competição entre programas é racionalmente avaliável. Um programa de pesquisa é progressivo se continuamente faz predições novas, surpreendentes e arriscadas que são subsequentemente confirmadas empiricamente, expandindo o conhecimento genuíno; e é degenerativo, se apenas acomoda fatos já conhecidos, através de ajustes *ad hoc* retrospectivos, não gerando conhecimento novo, mas apenas salvando os fenômenos. A escolha racional é clara para ele, a saber, sempre preferir o programa progressivo.

Mas, conforme dissemos, a demarcação entre o programa que é temporariamente anômalo e permanentemente degenerativo não é objetiva e temporal. O programa newtoniano enfrentou anomalias sérias (órbita de Mercúrio, órbita da Lua), mas não era degenerativo, pois eventualmente recuperou a fertilidade. Quanto tempo devemos esperar antes de abandonar um programa anômalo? Não há resposta objetiva independente do contexto. Isso deixa um espaço essencial aberto para um julgamento subjetivo ou social, enfraquecendo alegações de racionalidade algorítmica completa.

Laudan (1977) propôs um modelo mais nuançado baseado na solução de problemas. As teorias resolvem problemas empíricos explicando fenômenos e problemas conceituais indicando inconsistências internas, incompatibilidades com teorias de fundo. A mudança científica ocorre quando uma teoria nova

resolve mais problemas que a antiga, ou resolve problemas mais importantes, ou gera menos problemas conceituais próprios novos.

Isso permite *trade-offs* sofisticados. Por exemplo, uma teoria pode ser empiricamente menos precisa que a rival, mas conceitualmente mais coerente. Assim, os cientistas ponderam diferentes tipos de problemas diferentemente, dependendo do contexto e dos valores envolvidos. A mudança científica não é binária (progressivo x degenerativo) mas multidimensional e gradual. Mas isso ainda requer julgamento sobre quais problemas são importantes, quais soluções são adequadas, e como ponderar problemas empíricos e conceituais. Pesquisadores razoáveis discordam legitimamente sobre essas questões. A mudança científica não é puramente racional no sentido de logicamente compelida por evidência e regras universais. Envolve valores epistêmicos, julgamentos contextuais, até fatores sociológicos. Mas isso não a torna irracional; apenas reconhece que a racionalidade científica é mais rica e complexa que a demonstração dedutiva formal.

Minha posição é a de que Kuhn (1962) estava parcialmente correto, mas exagerou. A mudança científica não é tão cumulativa quanto a tradição empirista ingênua afirma, nem tão revolucionária e descontínua quanto Kuhn (1962) às vezes sugere. Há elementos de ambos, rupturas e continuidades entrelaçadas. Algumas mudanças são genuinamente revolucionárias, como a relatividade especial contra a mecânica newtoniana sobre tempo absoluto, ou a mecânica quântica contra o determinismo clássico. Outras são de fato refinamentos cumulativos, como a genética molecular integrando e expandindo a genética mendeliana. A maioria combina elementos de ambos.

A incomensurabilidade existe mas é parcial e superável. Conforme vimos, as teorias sucessivas frequentemente compartilham uma estrutura matemática, mesmo quando as ontologias diferem radicalmente. As equações de Maxwell sobreviveram à eliminação do éter. As leis de conservação persistem através de revoluções. Há mais continuidade estrutural do que Kuhn (1962) reconheceu. A mudança científica é mais racional do que Kuhn (1962) sugeria, mas menos algoritmicamente determinada que Popper (1934/1959) exigia. Há boas razões para mudanças de paradigmas, mas as razões não nos compelem univocamente. A racionalidade científica é um julgamento informado por evidências, valores

epistêmicos e consequências, não uma dedução automática de regras universais.

6. Unidade e Diversidade das Ciências

Até aqui discutimos problemas comuns a todas as ciências. Indução, demarcação, realismo e mudança científica são questões que qualquer filosofia da ciência deve enfrentar, independentemente de tratar de física, biologia, antropologia ou economia. Mas há uma tensão fundamental estruturando este livro, a tensão entre unidade e diversidade das ciências.

Há uma longa tradição, associada ao positivismo lógico e ao programa de uma "ciência unificada", que busca unidade metodológica e ontológica. Todas as ciências, por mais superficialmente diferentes que sejam, compartilham um método comum e podem ser unificadas em uma estrutura conceitual única. A física seria a ciência fundamental, e as outras ciências seriam redutíveis a ela através de cadeias de reduções interteóricas. A biologia reduz-se à química molecular, que reduz-se à física de partículas; a psicologia reduz-se à neurobiologia, que reduz-se à bioquímica, que reduz-se à física; as ciências sociais eventualmente reduzir-se-ão à psicologia individual, completando a pirâmide reducionista.

Essa visão possui um apelo considerável. Se todas as ciências compartilham um método comum, podemos avaliá-las por critérios universais. Se todas são ontologicamente redutíveis à física, há uma imagem metafisicamente unificada do mundo sem lacunas ontológicas misteriosas. A unidade promete coerência intelectual e poder explicativo integrado.

Mas essa visão enfrenta problemas profundos. A redução interteórica raramente é realizada satisfatoriamente. Conceitos biológicos como "gene", "seleção natural", "função adaptativa" não parecem redutíveis sem resíduo a química e física. Conceitos psicológicos como "crença", "desejo", "intencionalidade" resistem teimosamente à redução neurobiológica. Conceitos sociais como "instituição", "norma", "poder" parecem irredutíveis a fatos sobre indivíduos. Além disso, diferentes ciências usam metodologias profundamente diferentes. A física usa experimentos controlados rigorosamente e matemática sofisticada. A antropologia usa observação participante prolongada e interpretação hermenêutica de

significados culturais. A astronomia é puramente observacional, não experimental. A História reconstrói eventos únicos e irrepetíveis. Tentar impor uma metodologia única sobre essa diversidade distorce as práticas reais.

A irredutibilidade dos conceitos biológicos, psicológicos e sociais conecta-se com o argumento da múltipla realizabilidade de Hilary Putnam (1975), i.e., um mesmo estado mental ou biológico pode ser realizado por diferentes substratos físicos, o que implica que as leis do nível superior não são dedutíveis das leis do nível inferior. Isso não prova que a redução é impossível em princípio, mas mostra que os programas redutivistas simples enfrentam um obstáculo estrutural que vai além de dificuldades técnicas contingentes.

Uma tradição alternativa, associada a Wilhelm Dilthey (1883/2010) e à hermenêutica filosófica, enfatiza uma certa diversidade irredutível. Dilthey (1883/2010) distinguiu *Naturwissenschaften* (ciências naturais) e *Geisteswissenschaften* (ciências do espírito/humanas). As primeiras buscam explicação causal através de leis gerais. As segundas buscam compreensão interpretativa. A física explica os fenômenos naturais através de leis universais. A história compreende as ações humanas através da interpretação contextual das intenções, significados e razões envolvidos. São empreendimentos radicalmente diferentes com métodos, objetivos e padrões de sucesso distintos. Nós, como filósofos, sabemos da diferença íntima entre causas e razões.

Essa visão também possui méritos, dado que reconhece a especificidade das ciências humanas, que lidam com seres dotados de intencionalidade, criadores de significados, que agem por razões e não apenas a partir de causas. Mas corre risco de criar um abismo intransponível entre as ciências, tornando impossível o diálogo interdisciplinar produtivo. Muitas disciplinas, como a psicologia, a economia, a linguística, a ciência cognitiva, parecem situar-se entre os extremos, compartilhando características de ambos lados.

Minha posição, que estrutura este livro, é um certo pluralismo pragmático. Reconheço que há problemas filosóficos comuns a todas as ciências, como os que discutimos neste capítulo. Afinal, todas enfrentam a indução, todas precisam se demarcar, todas enfrentam questões sobre realismo, todas passam por mudanças teóricas. A filosofia geral da ciência ilumina esses problemas compartilhados. Mas também penso que cada ciência possui certas peculiaridades irredutíveis, que exigem uma filosofia específica, atenta à sua natureza particu-

lar. A antropologia lida com a cultura e o significado, levantando questões sobre o relativismo que a física não enfrenta. A linguística enfrenta o debate sobre a relação entre linguagem e pensamento; problema este ignorado pela química. A matemática lida com a natureza de certos objetos abstratos, questão que a biologia não enfrenta. A ciência da informação enfrenta uma dualidade físico-semântica de informação que a filosofia do direito não aborda.

Cada disciplina tem questões distintivas não dedutíveis mecanicamente da filosofia geral da ciência. Elas exigem uma reflexão filosófica específica, atenta às práticas, conceitos e dilemas próprios. Por isso este livro dedica um capítulo a cada ciência particular que aborda. Não abordamos todas as ciências. Isso não era praticamente possível e nem pedagogicamente interessante, de modo que preferimos seguir um caminho em que apresentamos uma história e usamos o nosso próprio histórico para construir a jornada deste livro.

Espero que este capítulo seja útil para as discussões que estão por vir nos próximos capítulos. Os problemas gerais de filosofia da ciência fornecem um pano de fundo, um conjunto de questões, que todas as ciências enfrentam, ainda que diferentemente. Há uma relação dialética entre filosofia geral da ciência e filosofias das ciências específicas. A filosofia geral ilumina questões que as ciências particulares enfrentam, mas as ciências particulares, reciprocamente, desafiam e refinam a filosofia geral. A física quântica forçou a reavaliação da causalidade e do determinismo. A biologia evolutiva forçou-nos repensar a teleologia e o design. As ciências cognitivas estão forçando a reavaliação da representação e da consciência. E, finalmente, a ciência da computação revelou os limites da computabilidade que afetam o que pode ser conhecido.

7. Considerações Finais

Examinamos quatro problemas centrais da filosofia da ciência, a saber, indução, demarcação, realismo e mudança científica. Em cada caso, vimos que não há soluções simples ou consensuais definitivas. Cada problema gera debates que persistem há séculos e provavelmente continuarão indefinidamente. Isso pode parecer frustrante ou decepcionante. Por que a filosofia não resolve os problemas definitivamente, como a matemática deduz os teoremas? A resposta é que os problemas filosóficos profundos raramente admitem soluções

algorítmicas simples. São difíceis precisamente porque envolvem *trade-offs* entre objetos múltiplos e parcialmente incompatíveis. Queremos que a ciência seja justificável logicamente *a priori*, mas a indução não é dedutivamente válida. Queremos um critério claro e preciso de demarcação, mas todas as nossas tentativas falham de um modo ou de outro. Queremos que as teorias sejam verdadeiras e falem sobre a realidade profunda, mas a história mostra algumas substituições radicais. Queremos que a mudança seja puramente racional, mas ela envolve julgamentos e valores.

Mas isso não torna a filosofia inútil ou ociosa. Mesmo sem resoluções definitivas, a filosofia esclarece os problemas, expõe os pressupostos ocultos, avalia a força de argumentos, mapeia o espaço de possibilidades teóricas, de modo que os cientistas beneficiam-se de clareza filosófica, mesmo sem soluções definitivas. A filosofia não resolve problemas científicos, mas dissolve confusões conceituais que bloqueiam o progresso. Nesse sentido, a filosofia da ciência funciona como metaciência reflexiva, pois não produz resultados científicos, mas clarifica as condições sob as quais resultados científicos são possíveis e o que eles significam.

Saber que a indução carece de justificação lógica dedutiva, mas que a usamos legitimamente mesmo assim, é superior a assumir ingenuamente sua obviedade. Saber que a demarcação é genuinamente difícil e contextual é melhor que o dogmatismo sobre o que conta como ciência. Reconhecer que o realismo e o anti-realismo possuem argumentos desconcertantes é mais honesto que o partidarismo acrítico a um dos lados.

Os capítulos seguintes aplicarão essas lições gerais a ciências específicas, mostrando como os problemas gerais se manifestam particularmente e que certos problemas novos emergem. Não é uma aplicação mecânica de um molde geral, mas um diálogo entre geral e particular, onde cada um informa reciprocamente o outro. A tensão entre unidade e diversidade das ciências é, ela mesma, um problema filosófico que permeia o livro. A jornada pelas ciências particulares começa agora.

2

Filosofia da Antropologia

1. A Antropologia como Problema Filosófico

A antropologia surgiu no século XIX como o estudo científico da humanidade em sua diversidade cultural, biológica e evolutiva. Desde o início, porém, carregou questões filosóficas fundamentais que resistem a tratamentos meramente empíricos. Três ordens de problemas permeiam a prática antropológica, a saber, problemas metafísicos sobre a natureza da cultura e sua relação com a biologia, problemas epistêmicos sobre a possibilidade de conhecer culturas radicalmente diferentes, e problemas éticos sobre se podemos julgar moralmente práticas culturais distintas das nossas.

A tripartição metafísica/epistêmica/ética como estrutura organizadora da filosofia da antropologia é uma escolha metodológica que tem precedente explícito em Quine (1969), que distingue questões de ser (ontologia), de conhecimento (epistemologia) e de norma (ética) como domínios filosóficos relativamente autônomos. A força da estrutura aqui adotada está em mostrar que as três ordens de problemas não são independentes; as respostas a uma constrangem as respostas às outras, criando uma arquitetura filosófica coerente para a disciplina.

Esses problemas não são externos à antropologia, isto é, não são questões filosóficas que os antropólogos podem ignorar em seu trabalho empírico. Ao contrário, estruturam internamente a própria possibilidade da disciplina. Respostas diferentes conduzem a programas de pesquisa radicalmente distintos. Um antropólogo que assume que culturas são sistemas simbólicos autônomos fará etnografia de um jeito; outro que as vê como estratégias adaptativas evolutivas fará de outro. A filosofia não vem depois da ciência, como reflexão opcional. Ela

está lá desde o começo, moldando o que conta como pergunta, método e evidência.

A questão metafísica central seria o que é a cultura. É um sistema simbólico autônomo, irredutível a processos biológicos ou psicológicos individuais, conforme sustentaram Durkheim (1895/2007) e os funcionalistas? Ou é, em última instância, um epifenômeno de estratégias adaptativas evolutivas, conforme argumentam os sociobiólogos e psicólogos evolucionistas? Essa não é uma disputa meramente terminológica. Se a cultura é ontologicamente autônoma, então a antropologia precisa de métodos interpretativos próprios, irredutíveis às ciências naturais. Se a cultura é ontologicamente redutível, então a antropologia deveria, a princípio, ser absorvida por uma ciência unificada da natureza humana, baseada em biologia evolutiva e psicologia cognitiva.

A questão epistêmica é igualmente premente, a saber, como podemos conhecer culturas radicalmente diferentes da nossa? Se todo o conhecimento é mediado por categorias culturais, e se essas categorias variam entre culturas, então seria a compreensão intercultural possível? Ou estamos condenados a projetar nossas próprias categorias sobre os outros, vendo apenas reflexos de nós mesmos? Esse problema atravessa a antropologia desde seus primórdios, ganhando formulações sofisticadas no debate entre relativismo e universalismo.

E a meta-reflexão ética sobre se podemos julgar moralmente as práticas de outras culturas. O relativismo cultural, defendido por Boas (1940), Benedict (1934) e Herskovits (1972), sustenta que os valores morais são internos aos sistemas culturais e que não há padrão transcultural de avaliação. O universalismo moral, defendido por filósofos como Nussbaum (2000, 2011), sustenta que existem padrões morais válidos para todos os seres humanos, e que podemos e devemos julgar as práticas que violam esses padrões. Não é questão meramente acadêmica, pois tem consequências diretas para a política internacional, para as intervenções humanitárias e para os direitos humanos de um modo geral.

Meu objetivo neste capítulo é articular com precisão os problemas filosóficos subjacentes à antropologia, mapear as principais posições teóricas e seus argumentos, avaliar os méritos e limites de cada posição, e defender uma posição que, embora não resolva todos os dilemas, ofereça um caminho produtivo para a antropologia contemporânea.

A estratégia será apresentar cada problema separadamente, reconhecendo que estão interconectados; a resposta a um constrange as respostas aos outros. Se adotarmos uma ontologia cultural forte, a saber, se tomamos a cultura como um sistema simbólico autônomo, isso torna mais difícil defender o universalismo epistêmico e mais fácil defender o relativismo ético. Inversamente, se adotarmos um naturalismo evolutivo forte que toma a cultura como uma estratégia adaptativa, isso facilitaria a defesa do universalismo, mas dificultaria o reconhecimento da diferença radical entre as culturas. Essas interconexões mostram que não podemos escolher posições isoladamente. Precisamos de uma visão coerente que responda aos três problemas simultaneamente.

Vou argumentar que a antropologia enfrenta um dilema estrutural, não facilmente solucionável, entre dois imperativos conflitantes. Por um lado, ela precisa de algum universalismo, pois se não houvesse invariantes culturais humanos, a antropologia não seria uma ciência comparativa, mas mera descrição de particularidades incomparáveis. Por outro lado, precisa de algum particularismo, dado que sem reconhecimento da diferença, perderia sua função crítica de desnaturalizar nossas próprias categorias. Esse dilema é parte de sua estrutura constitutiva. As diferentes escolas antropológicas podem ser vistas como tentativas de resolver ou mitigar esse dilema, privilegiando um dos polos. Nenhuma solução é plenamente satisfatória, contudo algumas são mais defensáveis que outras. Especificamente, defenderei que uma forma de pluralismo pragmático, inspirada na filosofia de Putnam (1981) e em antropólogos como Rabinow (1986), oferece o melhor compromisso disponível, embora não elimine as tensões.

Este capítulo é o primeiro de uma jornada que nos levará da antropologia à linguística, desta à ciência da informação, depois às instituições de memória e exposição da memória, e finalmente às tecnologias que projetam o futuro. A antropologia é o ponto de partida porque pergunta o que é esse ser que cria cultura, que dá significado ao mundo, que constrói instituições e tecnologias. As respostas que damos aqui ecoarão em todos os capítulos seguintes.

Se a cultura é irredutível, então arquivos, bibliotecas e museus não são meros depósitos de objetos, mas sistemas de significado que merecem interpretação. Se há invariantes ou universais culturais humanos, então talvez possamos falar de direitos humanos no turismo, de justiça algorítmica na computação, de

dignidade na robótica. O pluralismo que defenderei aqui não é uma posição isolada; é a primeira manifestação de uma abordagem que reaparecerá em diferentes formas ao longo do livro, mais com uma intenção pragmática de ser pedagógico e apresentar diversas visões divergentes do que algo fundado em minhas opiniões pessoais.

2. A Natureza da Cultura

A questão metafísica fundamental da antropologia é sobre a natureza da cultura. Essa pergunta, aparentemente simples, carrega implicações profundas para como a disciplina se concebe e se pratica. Podemos identificar quatro respostas principais na história da antropologia, a saber, o idealismo cultural, o materialismo cultural, o naturalismo evolutivo e o emergentismo. Examinarei cada posição, seus argumentos principais e suas dificuldades.

O idealismo cultural, associado a Geertz (1973), Schneider e à tradição interpretativista, sustenta que a cultura é essencialmente um sistema de significados compartilhados, transmitidos simbolicamente. Para Geertz (1973, p. 5) , "o homem é um animal amarrado a teias de significados que ele mesmo teceu". A cultura não é o comportamento, mas o sistema de símbolos e interpretações que torna o comportamento inteligível. Conforme dissemos, essa visão tem raízes filosóficas na hermenêutica de Dilthey (1883/2010) e na fenomenologia de Husserl (1913/2006), que distinguiram as ciências do espírito das ciências da natureza; enquanto estas explicam causalmente, aquelas compreendem interpretativamente por meio de razões. A força do idealismo cultural está em capturar o que parece distintivo da antropologia, o seu foco em significados e razões, não em mecanismos causais.

A distinção diltheyana entre Erklären (explicação causal das ciências naturais) e Verstehen (compreensão interpretativa das ciências do espírito) é o fundamento filosófico do programa interpretativista de Geertz (1973). Essa distinção foi posteriormente radicalizada por Winch (1958), que argumentou que as ciências sociais não apenas usam métodos diferentes das ciências naturais, mas têm um objeto categorialmente distinto, pois enquanto as ciências naturais lidam com regularidades causais, as ciências sociais lidam com regras, razões e

significados que constituem, e não meramente acompanham, os fenômenos sociais.

A crítica de Geertz (1973) ao behaviorismo e ao funcionalismo é certeira. Você não entende por que alguém pisca o olho apenas observando o movimento da pálpebra. Você precisa saber se é uma piscada involuntária, uma piscadela cúmplice, ou uma paródia de uma piscadela. O significado não está no comportamento físico, mas na interpretação culturalmente situada.

O idealismo tem implicações metodológicas claras. Ele nos diz que a antropologia deve ser interpretativa, não explicativa no sentido das ciências naturais. Seu objetivo é "ler" a cultura como um texto, desvelando as camadas de significado que tornam a ação social inteligível. Isso não significa que as explicações causais sejam irrelevantes, mas que elas são secundárias, i.e., antes de explicar por que algo acontece, precisamos entender o que significa.

A metáfora da cultura como texto, central no programa de Geertz (1973), tem uma história filosófica que remonta à hermenêutica de Ricoeur, o texto é uma unidade de significado que se "distancia" da intenção original do autor e do contexto de enunciação, ganhando autonomia semântica. Para Ricoeur, isso torna a interpretação não uma recuperação da intenção original, mas uma produção de significado em diálogo com o texto. Aplicado à antropologia, isso implica que a etnografia não recupera o significado "autêntico" de uma cultura, mas produz interpretações em diálogo com ela, o que tem implicações diretas para a questão da objetividade etnográfica.

Mas o idealismo enfrenta problemas sérios. Primeiro, tende a tratar as culturas como totalidades coerentes e integradas, subestimando as suas contradições internas, as disputas de poder e a agência individual. Se a cultura é um texto, quem é o autor? Quem decide o que significa? Geertz (1973) fala de "sistemas de significados compartilhados", mas o compartilhamento é sempre parcial e contestado. Isso nos leva à dificuldade em explicar a mudança cultural. Se as culturas são sistemas simbólicos autônomos, de onde vêm as transformações e, ainda, como explicar as regularidades transculturais? O idealismo tende a privilegiar a diferença sobre a semelhança, o que é filosófica e politicamente problemático.

A crítica ao idealismo cultural é desenvolvida sistematicamente por Sherry Ortner (1984) sob o conceito de "prática", em que os atores sociais não são

meros portadores de símbolos, mas agentes que reproduzem e transformam estruturas através de práticas situadas. A teoria da prática, desenvolvida também por Bourdieu (1977) e Giddens, constitui uma resposta ao dualismo entre idealismo cultural e determinismo estrutural, oferecendo um quadro em que estrutura e agência são mutuamente constitutivas.

Por sua vez, o materialismo cultural, associado a Marvin Harris (1979) e Leslie White, sustenta que a cultura é primariamente uma adaptação aos constrangimentos materiais do ambiente. A infraestrutura material de tecnologia, a economia e a demografia determinam a superestrutura ideológica dos símbolos, dos valores e das ideias. As mudanças culturais, nessa teoria, são explicadas por meio de mudanças nas condições materiais. Essa visão tem raízes no materialismo histórico de Marx (1867/1996), mas os antropólogos materialistas desenvolveram suas próprias versões. Harris (1979) propôs o "materialismo cultural" como estratégia de pesquisa, ao explicar os padrões culturais por meio de custos e benefícios materiais. Por exemplo, a proibição hindu de comer carne bovina não seria primariamente um valor religioso, mas uma adaptação ecológica, dado que os bois são mais valiosos vivos do que mortos, por serem animais de tração, de produção de leite e, inclusive, de estrume.

A tese de Harris (1979) sobre a vaca sagrada indiana tornou-se um caso emblemático de debate entre o materialismo cultural e o idealismo. O contra-argumento idealista, desenvolvido por Marshall Sahlins (1976), é que a "utilidade" da vaca é ela mesma culturalmente constituída; o que conta como recurso valioso, e como se avalia o custo-benefício, depende de uma lógica simbólica prévia que o materialismo cultural pressupõe sem explicar. Para Sahlins (1976), a cultura não é uma adaptação ao mundo material; ela é a forma pela qual o mundo material é constituído como tal.

A força do materialismo está em evitar o idealismo e conectar a cultura a processos materiais reais. Explica muitas práticas culturais que parecem arbitrárias ou irracionais quando vistas apenas simbolicamente. Mostra que os símbolos não flutuam livremente, mas respondem a pressões seletivas. Talvez seu maior problema seja tender ao reducionismo, quando nem todas as práticas culturais são explicáveis por vantagens materiais. Isso subestima a autonomia relativa da cultura, de que, uma vez estabelecidas, as práticas culturais desenvolvem uma lógica própria, que pode persistir mesmo quando as condições

materiais mudam. Além disso, o materialismo tem uma séria dificuldade em explicar a diversidade, pois se a cultura fosse uma adaptação ao ambiente material, por que as sociedades em ambientes similares desenvolveriam culturas tão diferentes?

Outra opção teórica seria o naturalismo evolutivo, associado à sociobiologia de Wilson (1975), à psicologia evolucionista de Tooby & Cosmides (1992), e à memética de Dawkins (1976). Ela sustenta que a cultura deve ser explicada em termos de processos darwinianos. Para os psicólogos evolucionistas, a mente humana consiste em módulos cognitivos especializados, selecionados durante o Pleistoceno para resolver problemas adaptativos recorrentes, como as capacidades de detectar trapaça, escolher parceiros, evitar predadores. A cultura é o resultado da interação desses módulos universais com ambientes locais. Alternativamente, para os memeticistas, a cultura consiste em unidades de informação, chamadas "memes", que se replicam através de cérebros humanos, sujeitos à variação, seleção e herança, análogos aos genes.

O naturalismo tem a virtude da unificação, pois promete explicar a cultura usando os mesmos princípios que explicam a vida biológica, evitando dualismos. Além disso, pode explicar tanto as propriedades culturais universais quanto a variação, pois nos diz que as primeiras refletem a arquitetura cognitiva universal, a última reflete os ambientes locais e as histórias contingentes das sociedades.

Entretanto, o naturalismo também tem os seus problemas. Ele comete frequentemente o erro de adaptar histórias de modo *ad hoc*. Para cada traço cultural, inventa-se uma história evolutiva plausível, mas não testável. Por exemplo, por que os humanos são monogâmicos? Porque isso aumenta o investimento paterno. Por que são polígamos? Porque isso aumenta a reprodução masculina. A teoria explica tudo, logo não explica nada. E, nesse sentido, ela tem também uma dificuldade em explicar a transmissão cultural, pois, diferentemente de genes, que são copiados com alta fidelidade, as ideias culturais são transformadas criativamente na transmissão. Assim, a analogia gene-meme é enganosa. De fato, o naturalismo parece cometer a falácia naturalista, ao mover-se indevidamente do ser para o dever, de fatos evolutivos para normas culturais. Mesmo se certas tendências foram selecionadas, isso não as torna justificadas.

A crítica da psicologia evolucionista por construção de narrativas *ad hoc* não testáveis é sistematizada por David Buller (2005) sob o rótulo de "teoria evolucionária da mente padrão" (s*tandard evolutionary psychology*), que nos diz que os modelos adaptativos *post-hoc* não distinguem adaptações funcionais de subprodutos evolutivos ou de características neutras, e frequentemente pressupõem o que pretendem explicar, ao escolher quais problemas do Pleistoceno são considerados relevantes. A alternativa metodológica proposta por Buller (2005) é a análise de adaptação comparativa rigorosa, análoga à sistemática filogenética em biologia.

O que fazer então? Qual teoria escolher? Uma alternativa às posições anteriores é o emergentismo. A ideia central é que a cultura é causalmente dependente de processos biológicos e materiais, mas constitui um nível emergente com propriedades próprias, não redutíveis ao nível do qual ela emergiu. A emergência, aqui, significa que as propriedades culturais, de nível superior, surgem das interações entre os elementos do nível inferior, como cérebros, corpos e ambientes, de tal forma que não podem ser explicadas apenas em termos desses elementos. O exemplo clássico é a água. Moléculas de H_2O individuais não são líquidas; a liquidez emerge quando muitas moléculas interagem. Similarmente, neurônios individuais não têm crenças ou desejos; crenças e desejos emergem de interações neurais em larga escala.

Há uma distinção importante na filosofia entre emergência fraca e emergência forte. Emergência fraca significa que as propriedades de nível superior são explicáveis a princípio a partir de propriedades de nível inferior, mas a explicação é complexa e não redutível a leis simples. A emergência forte implica que as propriedades de nível superior são ontologicamente irredutíveis, i.e., não podem ser explicadas completamente a partir de propriedades de nível inferior, mesmo a princípio. Para a cultura, a maioria dos emergentistas defende a emergência fraca de a cultura depender dos cérebros e das interações sociais, mas tem suas próprias regularidades que merecem um estudo autônomo.

A distinção entre emergência fraca e forte tem implicações diretas para a questão da autonomia metodológica da antropologia. Se a cultura exibe apenas emergência fraca, então explicações em termos de nível inferior são, em princípio, possíveis, e a autonomia da antropologia seria apenas pragmática (as explicações de nível inferior são demasiado complexas para serem práticas). Se

existe a emergência forte, a autonomia seria ontológica, pois haveria propriedades culturais que simplesmente não seriam redutíveis a propriedades biológicas ou materiais. A maior parte das posições contemporâneas opta pela emergência fraca combinada com a autonomia pragmática, evitando o dualismo substancial sem eliminar o nível cultural como objeto legítimo de investigação independente.

O emergentismo tem vantagens importantes com relação às outras teorias, como preservar a autonomia relativa dos significados (como o idealismo) sem rejeitar a base material (como o materialismo e o naturalismo). Além disso, ele também reconhece que a cultura é real e causalmente eficaz, e não misteriosa ou sobrenatural, oferecendo uma ontologia compatível com o naturalismo científico, sem cair no reducionismo.

Porém nem tudo são flores quando somos emergentistas. Uma das dificuldades diz respeito à própria palavra "emergência", que pode esconder mais do que revelar. Sem teorias específicas sobre como as propriedades culturais emergem das interações materiais, o termo vira meramente uma palavra mágica. E, ainda, a emergência fraca pode não ser suficiente para justificar a autonomia metodológica da antropologia, dado que, se as explicações em termos de nível inferior são possíveis, ao menos a princípio, por que não buscá-las? E, por sua vez, a emergência forte parece misteriosa e difícil de conciliar com o naturalismo. Qualquer escolha aqui é extremamente problemática.

Onde isso nos deixa? Temos várias teorias, cada uma com vantagens e desvantagens. O idealismo captura a autonomia dos significados, mas enfrenta problemas com o dualismo, com a universalidade e com a mudança. O materialismo oferece explicações causais robustas, mas reduz os significados a epifenômenos. O naturalismo unifica a cultura e a natureza, mas corre o risco de criar histórias não testáveis e de acabar aceitando uma falácia naturalista. O emergentismo parece promissor, mas requer uma maior precisão conceitual e desenvolvimento teórico.

Proponho uma posição intermediária, que chamarei de realismo cultural modesto. Ele se compromete com a ideia de que a cultura é real, e não meramente uma ficção conveniente, tendo uma existência objetiva, entretanto emergente de substratos biológicos e materiais, constituindo um nível próprio de explicações não redutíveis ao nível inferior. Dessa forma, as explicações culturais não seriam redutíveis a explicações biológicas ou psicológicas, embora

sejam consistentes com elas. E também não seriam fixas, porque a própria cultura não é fixa, mas muda através de processos que envolvem tanto fatores materiais quanto simbólicos.

Essa posição evita os extremos do idealismo, que trata cultura como puramente simbólica e autônoma, do materialismo, que a reduz a um epifenômeno, e do naturalismo, que a biologiza. Reconhece que a antropologia precisa de métodos interpretativos para os significados, causais para os processos materiais e evolutivos para as escalas temporais longas. Nenhum método sozinho é suficiente; a disciplina é intrinsecamente pluralista.

O "realismo cultural modesto" aqui proposto é compatível com o "naturalismo sem reducionismo" de Hilary Putnam (1975), no qual o compromisso com a realidade objetiva de entidades de nível superior (estados mentais, fatos sociais, propriedades culturais) não exige sua redução a entidades físicas, mas exige sua superveniência sobre elas. A superveniência garante a consistência com o naturalismo sem implicar identidade, pois não pode haver diferença cultural sem alguma diferença material, mas diferenças materiais distintas podem realizar a mesma propriedade cultural.

Essa conclusão metafísica terá implicações para os capítulos seguintes. Se a cultura é real, múltipla, emergente e histórica, então os produtos culturais (linguagem, informação, documentos, artefatos) herdam essas propriedades. A linguagem não é apenas um sistema formal abstrato, mas uma prática social emergente. A informação não é apenas bits, mas significado situado. Os documentos não são apenas objetos físicos, mas entidades com dimensões materiais e simbólicas. O pluralismo ontológico que defendo aqui ecoará em toda a nossa jornada.

3. Os Limites do Conhecimento Antropológico

Mesmo que resolvamos a questão metafísica sobre o que é a cultura, enfrentaremos uma questão epistêmica igualmente difícil, que é a de como podemos conhecer culturas radicalmente diferentes da nossa. Dado que toda compreensão é mediada por categorias conceituais, e dado que essas categorias variam entre culturas, como a compreensão intercultural é possível? Esse pro-

blema foi dramatizado pelo debate entre relativismo e universalismo na antropologia cognitiva e na filosofia da linguagem.

O relativismo epistêmico radical, inspirado na hipótese de Sapir-Whorf e desenvolvido por filósofos como Peter Winch (1964), sustenta que os esquemas conceituais são incomensuráveis, de modo que culturas diferentes habitam mundos diferentes. Para Winch (1964), seguindo Wittgenstein (1953/2009), entender uma cultura é participar de suas formas de vida, e as formas de vida são autônomas, sem critérios externos de racionalidade. O que conta como razão, evidência ou verdade varia de cultura para cultura. Portanto, o antropólogo não pode julgar as crenças de uma cultura usando os padrões de outra; ele só pode descrever internamente como aquela cultura funciona.

Winch (1964) ilustra seu argumento com o exemplo dos Azande, povo do Sudão estudado por Evans-Pritchard (1937). Os Azande consultam oráculos de veneno de galinha para tomar decisões importantes. O veneno é administrado a galinhas, e a sobrevivência ou morte da ave indica a resposta do oráculo. Para um observador ocidental, isso parece superstição irracional. Mas Winch (1964) argumenta que, dentro da forma de vida Azande, o oráculo é perfeitamente racional. Ele faz parte de um sistema coerente de crenças sobre bruxaria, magia e causalidade. Não podemos dizer que os Azande estão errados; eles têm padrões de racionalidade diferentes dos nossos, igualmente válidos dentro de sua forma de vida.

O relativismo radical tem o mérito de nos alertar contra o etnocentrismo. Muitas vezes, julgamos outras culturas com base em nossos próprios padrões, sem perceber que esses padrões não são universais. O relativismo nos força a levar a sério a possibilidade de que outras formas de vida sejam genuinamente diferentes, e não versões imperfeitas da nossa.

Entretanto, sua principal objeção advém de Davidson (1973) e Putnam (1981). Davidson (1973) argumentou que a incomensurabilidade é um pseudo-conceito. Para que possamos sequer identificar um esquema conceitual como diferente do nosso, já precisamos ter traduzido suficientemente para reconhecer a diferença, ao menos para reconhecer que se trata de outro esquema conceitual do mesmo tipo (o tipo cultural). Se não podemos traduzir absolutamente nada, não temos razão alguma para dizer que existe outro esquema conceitual; para nós, tal esquema incomensurável seria indistinguível de ruído ou loucura.

Portanto, esquemas conceituais totalmente incomensuráveis são impossíveis. Dessa forma, ou há tradução suficiente para reconhecer a diferença, caso em que há comensurabilidade parcial, ou não há base para dizer que há outro esquema.

O argumento davidsoniano contra a incomensurabilidade tem uma estrutura formal importante; ele é uma aplicação do princípio da caridade interpretativa à escala intercultural. Para que possamos atribuir crenças e significados a falantes de outra língua ou cultura, devemos pressupor um grau mínimo de racionalidade e de verdade nas suas crenças, caso contrário, não há base para a interpretação. A incomensurabilidade total seria, portanto, epistemicamente invisível, dado que não poderíamos detectá-la nem afirmá-la com fundamento.

Putnam (1981) ofereceu um outro argumento. O relativista radical comete uma falácia de inferência. Do fato de que toda compreensão é mediada por categorias conceituais, não se segue que não há verdade objetiva ou que diferentes esquemas são igualmente válidos. Categorias conceituais são ferramentas para lidar com o mundo, e algumas ferramentas são melhores que outras para certos propósitos. Que diferentes culturas usem ferramentas diferentes não mostra que o mundo não existe independentemente delas, mas apenas que há múltiplas maneiras de descrevê-lo. Putnam (1981) distingue o "relativismo conceitual" de reconhecer que há múltiplos esquemas conceituais do "relativismo sobre a verdade" de negar que há fatos objetivos. O primeiro é trivialmente verdadeiro; o segundo é incoerente.

No extremo oposto, encontramos o universalismo epistêmico. Horton (1967), em um debate famoso com Winch (1964), argumentou que as crenças dos Azande sobre oráculos, embora funcionem dentro de sua forma de vida, são objetivamente falsas. Galinhas não detectam bruxaria. Os Azande estão simplesmente errados, e podemos mostrar isso através de experimentos que eles próprios aceitariam como válidos. Horton (1967) sugere que todas as culturas compartilham padrões de racionalidade mínimos, que são usados para buscar explicações, para avaliar evidências, para reconciliar crenças inconsistentes. A diferença é que a comunidade científica desenvolveu instituições, como a crítica acadêmica e a experimentação sistemática, que tornam essas práticas mais eficazes.

Uma das vantagens do universalismo é que ele torna o conhecimento antropológico possível, porque, se os antropólogos podem aprender línguas exóticas e entender crenças aparentemente estranhas, isso mostra que há uma racionalidade compartilhada que é suficiente para a tradução. O que seria consistente com as descobertas das ciências cognitivas sobre propriedades universais humanas. E teria a grande vantagem de preservar a possibilidade da crítica intercultural.

A evidência empírica das ciências cognitivas em favor de algum universalismo é substancial. As pesquisas sobre cognição de cor, categorização biológica, intuições morais básicas e aquisição de linguagem sugerem que a mente humana não é uma "lousa em branco" que as culturas preenchem arbitrariamente, mas possui estruturas inatas que constrangem a variação cultural. O debate contemporâneo não é entre universalismo total e relativismo total, mas sobre quais aspectos da cognição são universais e quais são culturalmente variáveis.

Porém sua primeira desvantagens é que ele pode ser uma forma de etnocentrismo disfarçado. Ao postular padrões universais de racionalidade, corremos o risco de simplesmente universalizar os padrões da ciência ocidental. Mesmo que haja universais, a variação cultural é real e importante, e o universalismo pode negligenciá-la. E a universalidade dos padrões de racionalidade defendido pelo universalismo lhe daria uma certa dificuldade em explicar por que culturas diferentes chegam a crenças tão diferentes, se todas compartilham os mesmos padrões básicos de racionalidade.

Putnam (1981) desenvolveu uma posição intermediária que ele chamou de "realismo interno". A ideia central é que há múltiplos esquemas conceituais igualmente bons para descrever o mundo, mas nem todos os esquemas conceituais possíveis são bons. Os esquemas devem ser avaliados pragmaticamente, i.e., por meio de quão bem nos permitem prever, explicar, intervir. Culturas diferentes podem ter esquemas diferentes, ambos funcionais, sem que um seja objetivamente superior. Mas isso não implica que todos os esquemas estão em pé de igualdade; os esquemas que levam a predições sistematicamente falsas são piores.

Essa posição evita tanto o relativismo radical, que aceita todos os esquemas como igualmente válidos, quanto o universalismo ingênuo, que postula um único esquema correto. Dessa forma, ela nos diz que a escolha entre esquemas

conceituais depende de propósitos e contextos, mas não é arbitrária, dado que alguns esquemas são melhores para certos propósitos; outros, para propósitos diferentes.

Aplicado à antropologia, o realismo interno sugere que podemos avaliar crenças e práticas de outras culturas não por correspondência com uma verdade absoluta, mas por sua adequação a propósitos e contextos. O oráculo Azande pode ser inadequado para prever eventos naturais, mas adequado para resolver disputas sociais ou tomar decisões em condições de incerteza. Não precisamos escolher entre dizer que é "racional" ou "irracional" em termos absolutos; podemos avaliá-lo em contextos específicos.

A avaliação contextual das práticas de conhecimento de outras culturas converge com o pragmatismo de Richard Rorty (1979), porque em vez de perguntar se uma crença "corresponde à realidade", devemos perguntar se ela funciona, para que propósitos e em que contextos. Rorty (1979) aplicou essa perspectiva ao debate sobre racionalidade cultural, argumentando que a diferença entre ciência e magia não é uma diferença na capacidade de acessar a verdade, mas uma diferença nos propósitos que cada prática serve e nos critérios de sucesso que cada comunidade adota.

Essas questões epistêmicas têm implicações práticas diretas para a pesquisa antropológica. Como devemos conduzir o trabalho de campo? Como interpretar o que nossos informantes dizem? Como escrever etnografias? Uma posição pluralista sugere que diferentes métodos são apropriados para diferentes propósitos. Para entender sistemas de significado, precisamos de métodos interpretativos, como observação participante, entrevistas e análise de discurso. Para entender os processos causais, precisamos de métodos explicativos, como a comparação controlada e a análise histórica. Para entender escalas temporais longas, precisamos de métodos evolutivos, como os da arqueologia e da genética. Nenhum método sozinho é suficiente. A antropologia é intrinsecamente multi-metodológica. Isso é um reflexo da complexidade de seu objeto. Se a cultura é múltipla, tal como argumentei anteriormente, então ela tem necessidade de múltiplos métodos.

Essa conclusão conecta-se diretamente com o próximo capítulo, sobre linguística. Lá, enfrentaremos questões análogas. Como conhecemos línguas? Que métodos são apropriados? Julgamentos de gramaticalidade, dados de

corpus, experimentos psicolinguísticos; cada um tem suas forças e limitações. O pluralismo metodológico que emerge aqui reaparecerá lá e nos capítulos sobre ciência da informação, arquivologia e biblioteconomia, onde diferentes tipos de evidência (documentos, registros, comportamentos de usuários) exigirão diferentes abordagens.

4. Relativismo e Julgamento Moral

O problema ético mais controverso na antropologia talvez seja o de saber se podemos julgar moralmente práticas de outras culturas. Esse problema tem uma longa história na filosofia moral, mas ganhou urgência particular na antropologia do século XX, quando o relativismo cultural tornou-se quase um dogma disciplinar. A questão não é meramente acadêmica, mas tem implicações diretas para as políticas de direitos humanos, para as intervenções humanitárias, e para a crítica de práticas como mutilação genital feminina, casamento infantil, escravidão e canibalismo ritual.

O relativismo cultural, defendido por Boas (1940), Benedict (1934) e Herskovits (1972), sustenta que os valores morais são internos aos sistemas culturais e que não há padrão transcultural de avaliação. O que é certo ou errado varia de cultura para cultura, e nenhuma cultura tem autoridade para julgar a outra. Como disse Herskovits (1972), "os julgamentos são baseados na experiência, e a experiência é interpretada por cada indivíduo em termos de sua própria enculturação".

As motivações do relativismo são nobres. Ele surge como uma reação ao etnocentrismo e ao imperialismo ocidental, que usavam a sua suposta superioridade moral para justificar a conquista e a dominação. Se todas as culturas são igualmente válidas, então não há base para impor nossos valores aos outros. As práticas que nos parecem bárbaras quando isoladas podem fazer sentido dentro de seu contexto cultural. O infanticídio entre os Inuit, por exemplo, não reflete crueldade, mas uma resposta racional a terem recursos escassos em um ambiente ártico. Julgá-los sem entender esse contexto é etnocêntrico.

Ainda que uma boa motivação tenha trazido o relativismo à vida, filosoficamente, ele parece autocontraditório. Se não há padrões morais universais, então a própria prescrição relativista de que não devemos julgar outras culturas

não é universal, logo não é vinculante para nós. Se, contrariamente, for universal, então há pelo menos um padrão moral universal, o da tolerância, contradizendo o relativismo. Além dessa objeção lógica, há objeções práticas igualmente problemáticas. O relativismo nos proíbe de criticar práticas manifestamente injustas. Se a moralidade é relativa, não podemos condenar a escravidão nazista, o genocídio de Ruanda, ou a opressão talibã das mulheres. Isso parece um *reductio ad absurdum* da posição, dado que poucos antropólogos aceitariam consistentemente essa implicação.

A objeção da autocontradição ao relativismo moral é classicamente denominada "argumento da tolerância" (*tolerance argument*), segundo o qual o relativismo prescritivo que ordena não julgar outras culturas é ele mesmo uma norma moral que pretende validade universal, contradizendo a tese de que não há normas morais universais. A versão mais sofisticada desse argumento, desenvolvida por Wong (2006), é a de que o relativismo não precisa negar que haja padrões morais interculturalmente compartilhados; precisa apenas negar que haja um único sistema moral correto. Essa versão moderada escapa da autocontradição, mas perde o alcance crítico do relativismo original.

Mas as dificuldades não param por aí; uma das maiores é a de que ele pressupõe que as culturas sejam entidades homogêneas e consensuais, o que é empiricamente falso. Todas as culturas contêm dissenso interno, conflito, contra-cultura, opressão. Quem fala pela "cultura"? Os poderosos ou os oprimidos? Se as mulheres contestam a mutilação genital dentro de uma cultura que a pratica, elas expressam algo da cultura local, ou devemos silenciar suas vozes ao pensar sobre a cultura local? O relativista não tem uma resposta clara.

A crítica ao pressuposto da homogeneidade cultural é desenvolvida por Abu-Lughod (1991) sob o conceito de "culturas como argumento". Em vez de entidades coerentes, culturas são campos de disputa onde diferentes grupos sociais competem pela definição do que conta como norma legítima. Isso implica que o apelo ao relativismo cultural para silenciar críticas externas pode ser instrumentalizado pelos grupos dominantes para sufocar dissidências internas, invertendo a motivação original do relativismo, que era proteger grupos subalternos do imperialismo ocidental.

No extremo oposto, encontramos o universalismo moral, que sustenta que existem padrões morais válidos para todos os seres humanos, independente-

mente de sua cultura. Podemos e devemos julgar práticas que violam esses padrões, como a mutilação genital feminina, a escravidão, ou a tortura. Isso tem uma força intuitiva significativa para nós. Quando vemos sofrimento humano desnecessário, reconhecemos a injustiça, independentemente do contexto cultural. O universalismo fundamenta os direitos humanos e permite a crítica de práticas opressivas. Além disso, é internamente coerente e não auto-refutante.

Porém, tal como dissemos, ele pode ser uma forma de etnocentrismo disfarçado, sendo usado para intervenções imperialistas. Muitas vezes, o que se apresenta como valor universal são, na verdade, valores ocidentais particulares. A ênfase na autonomia individual, por exemplo, pode não ter o mesmo peso em culturas que privilegiam a comunidade ou a família.

Entre os extremos do relativismo absoluto e do universalismo rígido, encontramos posições intermediárias que tentam preservar intuições de ambos os lados. Sen (1999) e Walzer (1994), por exemplo, propõem um pluralismo de valores, ao nos dizer que há múltiplas concepções de bem humano, não redutíveis a uma métrica única, mas há também proibições mínimas universais (não escravizar, não torturar, não matar inocentes). Dentro dessa "zona mínima", práticas podem ser universalmente condenadas; fora dela, há espaço para variação legítima. Essa abordagem permite tanto a diversidade quanto a crítica limitada.

Outro caminho seria o procedimentalismo de, em vez de especificar os valores substantivos universais, especificar os procedimentos universais, como os da democracia deliberativa ou do diálogo intercultural. As práticas seriam justificadas se sobreviverem ao escrutínio, em condições de deliberação livre. Isso evitaria o imperialismo, pois quem julga não é um grupo que se julga moralmente superior, mas o grupo que se constitui a partir de um diálogo inclusivo. Contrariamente, as práticas que não sobreviverem ao diálogo inclusivo seriam consideradas injustificadas.

O procedimentalismo habermasiano enfrenta, no contexto intercultural, um problema de *bootstrapping*, pois as regras do diálogo racional (universalidade, ausência de coerção, igualdade de participação) são elas mesmas normas substantivas que não são culturalmente neutras. Seyla Benhabib (2002) reformula essa crítica como o problema do "paradoxo democrático", i.e., qualquer tentativa de fundamentar normas universais em procedimentos dialógicos pres-

supõe já a aceitação de normas que garantem a validade do procedimento, o que é uma forma de circularidade.

Appiah (2006) desenvolve uma posição que ele chama de "cosmopolitismo enraizado". Há obrigações universais para com a humanidade, mas também obrigações especiais para com grupos particulares, baseadas em história, identidade e outras relações. Aplicado à antropologia, isso sugere que podemos fazer julgamentos transculturais, mas devemos fazê-los com humildade, abertura e disposição para aprender com as perspectivas locais.

Consideremos dois exemplos concretos para testar essas posições. A Mutilação Genital Feminina (MGF) é praticada em várias culturas africanas e do Oriente Médio. Ela envolve a remoção parcial ou total dos órgãos genitais femininos. Para muitos no Ocidente, é uma violação grave dos direitos humanos, causando dor, sofrimento e problemas de saúde duradouros. Para algumas comunidades praticantes, é um rito de passagem necessário, que prepara meninas para a feminilidade e o casamento, e está profundamente entrelaçado com a identidade cultural e com valores de pureza.

Um universalista puro condenaria a prática sem hesitação, apelando a direitos humanos universais. Um relativista puro se recusaria a julgar, argumentando que devemos respeitar a integridade cultural. Mas ambas as posições parecem insatisfatórias. O universalista ignora que as mulheres nas comunidades praticantes podem ter perspectivas diferentes, e que intervenções externas podem ser contraproducentes. O relativista ignora que muitas mulheres dentro dessas comunidades contestam a prática, e que silenciá-las em nome do respeito cultural é outra forma de opressão.

Como o nosso pluralismo pragmático lidaria com isso? Buscando compreender o que significa para as mulheres que a vivenciam e quais são os discursos locais a favor e contra, para apoiar processos de mudança endógenos, fortalecendo vozes locais que contestam a prática, em vez de impor soluções externas. A ideia de fundo é que há danos objetivos causados às pessoas, como dor, sofrimento e riscos à saúde, que não dependeriam de interpretação cultural.

A estratégia de apoiar mudanças endógenas em vez de impor soluções externas tem fundamentação filosófica na abordagem das capacidades (*capabilities approach*) de Nussbaum (2000, 2011) e Sen (1999). Em vez de impor um

conjunto específico de valores culturais, a abordagem identifica capacidades humanas centrais (saúde corporal, integridade física, controle sobre o próprio ambiente) cuja proteção é condição de possibilidade para que as pessoas possam viver escolhas genuinamente suas. Isso permite a crítica de práticas que bloqueiam capacidades centrais sem impor um ideal específico de vida boa.

Em algumas comunidades tradicionais, meninas são casadas muito jovens, às vezes antes da puberdade. Para defensores de direitos humanos, isso é uma violação da autonomia e do bem-estar da criança. Para comunidades praticantes, pode ser uma forma de proteger meninas, garantir alianças familiares, ou cumprir mandatos religiosos. Novamente, nem o universalismo nem o relativismo puros são adequados. O pluralismo pragmático buscaria entender as funções sociais do casamento infantil na comunidade, os discursos locais a favor e contra, e as possibilidades de mudança negociada, apoiando iniciativas que empoderam meninas e famílias a fazer escolhas diferentes, sem impor soluções de fora.

Essas questões éticas têm implicações diretas para a prática antropológica. Como devemos nos posicionar em relação às comunidades que estudamos? Temos obrigações para além da pesquisa? Devemos intervir quando testemunhamos injustiças? E o que devemos pensar como injustiça?

Uma posição defensável é que os antropólogos têm obrigações especiais para com as comunidades que estudam, derivadas da relação de confiança estabelecida durante o trabalho de campo. Isso inclui o respeito pela autonomia local, mas também a responsabilidade de não ser cúmplice de injustiças ou danos objetivos às pessoas. Em casos de dano grave, pode ser eticamente exigível intervir, mesmo que isso signifique violar a neutralidade. Mas a intervenção deve ser feita com cuidado, buscando minimizar os danos e respeitar a agência local.

Essas questões ecoarão em capítulos posteriores. Na arquivologia, enfrentaremos dilemas que colocam em conflito o acesso a documentos sensíveis e a privacidade dos indivíduos. Na biblioteconomia, encontraremos dilemas sobre a neutralidade e a responsabilidade social. Na museologia, sobre repatriação de objetos e autoridade interpretativa. No turismo, sobre os impactos do turismo em comunidades locais. Na computação e na engenharia, sobre automação, vigilância e justiça algorítmica. O pluralismo pragmático que defendo aqui, que

reconhece valores múltiplos, contextos variados e a necessidade de julgamento prudencial, será uma ferramenta conceitual útil em todos esses domínios.

5. Pluralismo Pragmático como Resposta

Examinei três ordens de problemas filosóficos na antropologia. Sobre a natureza da cultura, vimos que nem idealismo, nem materialismo, nem naturalismo evolutivo, nem emergentismo são plenamente satisfatórios; propus um realismo cultural modesto que reconhece a cultura como real, múltipla, emergente e histórica. Sobre o conhecimento antropológico, vimos que nem relativismo radical nem universalismo forte funcionam; propus um pluralismo metodológico que integra diferentes tipos de evidência e reconhece a legitimidade de múltiplos quadros conceituais, avaliados pragmaticamente. Sobre a moralidade, vimos que nem relativismo absoluto nem universalismo rígido são defensáveis; então propus um universalismo mínimo processual que permite a crítica limitada sem imperialismo.

Há um padrão comum nessas respostas, a saber, a rejeição de posições extremas e adoção de posições intermediárias que tentam preservar intuições de ambos os lados. Isso pode parecer insatisfatório, mero ecletismo. Mas penso que reflete a estrutura do problema; a antropologia enfrenta dilemas genuínos, não meras confusões conceituais. O dilema estrutural é que a antropologia precisa de universalismo suficiente para ser uma ciência comparativa, mas também precisa de particularismo suficiente para cumprir sua função crítica de nos apresentar novas categorias e desnaturalizar nossas próprias. Essas exigências puxam em direções opostas. Nenhuma escola resolveu o dilema completamente.

Proponho que abandonemos a busca por uma solução definitiva e adotemos o que chamo de pluralismo pragmático, inspirado em Putnam (1981) e em antropólogos como Rabinow (1986). A ideia central é que não há um único quadro conceitual correto para a antropologia, mas múltiplos quadros igualmente legítimos, cada um com certas vantagens para certos propósitos. Para alguns propósitos, como estudar universais culturais e testar hipóteses evolutivas, um quadro naturalista é apropriado. Para outros, como entender os significados locais e capturar a experiência vivida, um quadro interpretativo é mais apropria-

do. Para ainda outros, como criticar injustiças ou intervir em conflitos, um quadro normativo é necessário. Esses quadros não são redutíveis uns aos outros, mas também não são incomensuráveis; eles se sobrepõem parcialmente e podem informar-se mutuamente.

O pluralismo pragmático evita tanto o fundacionismo de buscar um único fundamento correto quanto o relativismo de aceitar todos os quadros igualmente. A escolha de quadro conceitual depende dos propósitos práticos que temos e dos contextos de investigação em que estamos inseridos. Isso não torna a escolha arbitrária; alguns quadros são melhores que outros para certos propósitos, e isso pode ser avaliado pragmaticamente, por meio de análise de quais predições funcionam, quais intervenções têm sucesso, quais interpretações são fecundas.

Algumas implicações práticas seguem desse pluralismo. A transdisciplinaridade não é opcional, mas necessária; a antropologia precisa dialogar com a biologia evolutiva para entender restrições universais, com a psicologia cognitiva para entender processos mentais, com a filosofia moral para articular compromissos éticos, e com os estudos culturais para entender o poder e a representação. Nenhuma disciplina tem o quadro completo.

Os antropólogos devem explicitar seus compromissos metateóricos em vez de fingir neutralidade. Toda etnografia pressupõe ontologia, epistemologia e ética; melhor torná-las explícitas e justificáveis do que escondê-las. E, como nossas interpretações são falíveis, nossos julgamentos são situados e nossas intervenções têm consequências não intencionais, a humildade epistêmica e moral é essencial. Isso não implica paralisia relativista, mas cautela justificada.

O pluralismo pragmático não resolve os dilemas da antropologia, mas oferece uma forma de conviver produtivamente com eles. Uma antropologia madura abraça essas tensões em vez de tentar eliminá-las através de teorias radicais demais. O pluralismo pragmático aqui proposto tem afinidade com o “realismo sem absolutismo” de Bernstein (1983), que articula uma posição filosófica que reconhece a objetividade sem postular fundamentos absolutos e reconhece a diversidade sem cair no relativismo. Para Bernstein (1983), a alternativa ao fundacionismo e ao relativismo não é uma posição intermediária de compromisso, mas uma concepção diferente de racionalidade, que emerge de práticas

situadas, que aprende com o diálogo, e que avalia por critérios pragmáticos sem absolutizá-los.

6. Considerações Finais

Examinei três problemas filosóficos centrais para a antropologia, a saber, a natureza da cultura, a possibilidade do conhecimento intercultural, e a legitimidade do julgamento moral transcultural. Em cada caso, mostrei que posições extremas são insustentáveis, mas que posições intermediárias são defensáveis, embora não eliminem totalmente as tensões.

O dilema estrutural da antropologia entre o universalismo necessário para a comparabilidade e o particularismo necessário para a crítica não tem resolução definitiva. Diferentes programas de pesquisa privilegiam diferentes pólos, e isso não é um erro, mas um reflexo de propósitos práticos diferentes. O pluralismo pragmático defende a legitimidade de múltiplos quadros, avaliados por critérios pragmáticos e contextualizados.

Isso tem implicações para a transdisciplinaridade. A antropologia não pode ser autossuficiente; ela precisa dialogar com as ciências naturais para entender os universais biológicos, com as ciências sociais para entender os processos históricos e econômicos, e com a filosofia para articular seus compromissos ontológicos, epistêmicos e éticos. A filosofia da antropologia não é só uma reflexão externa sobre a disciplina, mas uma parte essencial da prática antropológica reflexiva. Todo antropólogo é, implicitamente, um filósofo da antropologia; melhor ser explicitamente consciente dos compromissos filosóficos que estruturam sua pesquisa do que ignorante deles.

Este capítulo estabeleceu fundamentos que serão importantes em todo este livro. A cultura, como vimos, é real, múltipla, emergente e histórica. O conhecimento intercultural é possível, mas mediado e parcial. Os julgamentos morais são inevitáveis, mas devem ser feitos com humildade e abertura ao diálogo.

No próximo capítulo, sobre Filosofia da Linguística, enfrentaremos questões análogas. O que é uma língua? Como a linguagem se relaciona com o pensamento e com a realidade? Existem universais linguísticos ou a diversidade é irredutível? Que responsabilidades éticas temos para com línguas ameaçadas? Veremos que os mesmos dilemas entre universalismo e particularismo, explica-

ção e interpretação, descrição e crítica, reaparecem. E veremos também que o pluralismo pragmático, agora sob o nome de “pluralismo metodológico”, oferece um caminho igualmente promissor.

A antropologia nos ensinou que a cultura é o meio pelo qual os humanos criam significado. A linguística nos ensinará que a linguagem é o veículo pelo qual esse significado é transmitido. Nosso caminho é longo, mas começa aqui, com a pergunta sobre o que somos nós, seres de cultura. As respostas que demos, provisórias, plurais e pragmáticas, nos acompanharão até o fim.

3

Filosofia da Linguística

1. Por que uma Filosofia da Linguística?

Quando um linguista analisa a estrutura sintática de uma sentença, quando um sociolinguista documenta variações dialetais de uma comunidade, quando um psicolinguista investiga como crianças adquirem sua língua materna, não estão apenas coletando dados empíricos ou aplicando modelos formais. Estão engajados em práticas que pressupõem respostas a questões filosóficas fundamentais. O que é uma língua? Como a linguagem se relaciona com o pensamento e com a realidade? Existem universais linguísticos ou a diversidade é irredutível? Que responsabilidades éticas temos para com línguas ameaçadas ou comunidades linguísticas marginalizadas?

A tripartição das perguntas filosóficas da linguística em ontológicas (o que é uma língua), epistemológicas (como conhecemos línguas) e éticas (que responsabilidades temos) é paralela à estrutura do capítulo anterior sobre a antropologia, reforçando a arquitetura filosófica do livro. Essa correspondência não é acidental, pois a linguística enfrenta os mesmos dilemas estruturais entre universalismo e particularismo, entre explicação e interpretação, entre descrição e crítica, agora aplicados ao objeto específico da linguagem.

Essas questões não podem ser respondidas por experimentos psicolinguísticos, análises sintáticas ou modelos computacionais. Elas exigem argumentação filosófica. A linguística, em todas as suas subdisciplinas, enfrenta desafios teóricos e práticos que não podem ser adequadamente compreendidos ou resolvidos sem fundamentação filosófica explícita. Desde questões sobre a natureza ontológica da língua até dilemas éticos sobre preservação linguística e justiça social, a linguística opera em terreno filosófico, muitas vezes sem reconhecê-lo ou examiná-lo criticamente.

Se no capítulo anterior perguntamos o que é a cultura, agora perguntamos como a cultura se expressa e se transmite. Através da linguagem. A linguagem é o medium da cultura. Não só porque falamos sobre o que somos, mas porque a própria estrutura do pensamento, suspeitam alguns, é moldada pela língua que falamos. Compreender a linguagem é, portanto, essencial para compreender a cultura. E compreender a ciência que estuda a linguagem, i.e., a linguística, é essencial para compreender como nós, humanos, nos compreendemos.

É importante distinguir duas disciplinas relacionadas, mas distintas. A filosofia da linguagem investiga questões sobre a natureza da linguagem, seu funcionamento, e sua relação com o pensamento e a realidade. Perguntas clássicas incluem "como as palavras adquirem significado?", "o que é uma proposição?", "como a referência funciona?". Essas questões são de natureza semântica, pragmática e lógica.

A filosofia da linguagem como disciplina autônoma emergiu sobretudo do programa analítico de Frege (1983) e Russell, que procuraram fundamentar a matemática na lógica e, ao fazê-lo, desenvolveram ferramentas de análise semântica de alcance muito mais amplo. A distinção fregeana entre sentido (*Sinn*) e referência (*Bedeutung*) permanece um dos instrumentos mais precisos disponíveis para analisar questões sobre significado, identidade e verdade em linguagem natural, e tem implicações diretas para debates linguísticos sobre semântica lexical e estrutura informacional.

A filosofia da linguística, por sua vez, investiga os pressupostos metateóricos da ciência linguística em si. O que os linguistas estudam quando estudam línguas? Que métodos são apropriados? Quais compromissos ontológicos e epistemológicos estão implícitos nas teorias linguísticas? As duas áreas se sobrepõem, mas mantêm focos distintos. A filosofia da linguagem pergunta sobre a linguagem; a filosofia da linguística pergunta sobre a ciência da linguagem.

Considere três tipos de perguntas sobre linguagem. Perguntas empíricas como quantas línguas são faladas no Brasil?, ou como funciona o sistema de casos do latim?, ou que padrões de *code-switching* ocorrem em comunidades bilíngues?, são respondidas pela linguística descritiva. Perguntas formais, como qual é a complexidade computacional de gramáticas transformacionais?, ou como formalizar relações semânticas em lógica modal?, são respondidas pela

lógica e pela linguística matemática. Perguntas filosóficas são respondidas pela argumentação conceitual. Línguas existem independentemente de falantes? A linguagem determina o pensamento ou apenas o influencia? Existe uma gramática universal inata ou as línguas são inteiramente aprendidas? Linguistas têm obrigações éticas de documentar línguas ameaçadas ou isso seria paternalismo?

É este terceiro conjunto de perguntas que constitui o domínio próprio da filosofia da linguística. Enquanto a linguística investiga como as línguas funcionam, como são adquiridas, como mudam historicamente, a filosofia investiga o que são essas línguas, se e por que importam, quais normas devem governar sua investigação, e que tipos de valores a pesquisa linguística expressa ou realiza.

Alguém poderia objetar que a linguística é uma ciência empírica, orientada a descrever e explicar fenômenos linguísticos. Por que necessitaria de reflexão filosófica abstrata? Não seria mais produtivo focar em coleta de dados, análise de corpus e construção de teorias explanatórias?

Essa objeção erra em três níveis fundamentais. Ela pressupõe falsamente que a prática linguística é teoricamente neutra, quando na verdade toda descrição e análise incorpora pressupostos filosóficos que podem e devem ser examinados; e ela ignora que muitos dos problemas mais urgentes enfrentados por linguistas são fundamentalmente filosóficos, disfarçados de questões técnicas, subestimando o potencial da reflexão filosófica para iluminar, questionar e aprimorar a pesquisa linguística.

Conforme dissemos, não existe prática linguística teoricamente neutra. Quando um sintaxista adota a teoria gerativa e postula estruturas sintáticas abstratas, está operacionalizando uma ontologia particular sobre o que são objetos linguísticos. Quando um sociolinguista documenta a variação dialetal e evita julgamentos sobre "correção", está aplicando uma epistemologia e uma ética particulares sobre a legitimidade linguística. Quando um psicolinguista investiga o processamento sentencial usando o tempo de reação, ele está pressupondo uma teoria sobre a relação entre competência linguística e performance.

Esses pressupostos filosóficos raramente são explicitados ou examinados criticamente na formação profissional. O resultado é que linguistas frequente-

mente operam com teorias implícitas, não articuladas, potencialmente contraditórias, que nunca foram submetidas ao escrutínio filosófico. Uma filosofia da linguística torna explícitos esses pressupostos, permitindo que sejam examinados, debatidos, refinados ou rejeitados.

Muitos problemas que aparecem como questões técnicas ou empíricas são, em sua essência, filosóficos. O debate sobre gramática universal e aquisição de linguagem não é meramente empírico. Envolve questões metafísicas sobre a natureza humana, questões epistemológicas sobre o que pode ser aprendido a partir da evidência linguística, e questões metodológicas sobre como testar hipóteses inatistas. Chomsky (1965, 1986) e seus críticos não discordam apenas sobre os dados empíricos, mas sobre qual tipo de explicação conta como adequada nessa área.

A hipótese Sapir-Whorf de que a linguagem influencia ou determina o pensamento levanta questões profundamente filosóficas. O que significa "influenciar o pensamento"? O pensamento é possível sem a linguagem? A linguagem pública advém de alguma linguagem privada? Como distinguir a influência linguística da influência cultural? Essas são questões que transcendem os estudos experimentais, traçando fronteiras conceituais que fundamentam os próprios estudos experimentais.

A versão forte da hipótese Sapir-Whorf (determinismo linguístico) foi amplamente rejeitada, mas versões moderadas (relativismo linguístico) continuam sendo investigadas empiricamente. Os experimentos de Boroditsky (2001) sobre a influência de sistemas de cores, direções espaciais e gramática de gênero sobre a cognição sugerem efeitos reais, mas modestos e reversíveis, da língua sobre o pensamento. A questão filosófica relevante é se esses efeitos constituem influência sobre o pensamento ou apenas sobre as tarefas de verbalização, distinção que os próprios experimentos não resolvem de forma definitiva.

A questão sobre quando dois dialetos contam como línguas distintas ou se são variedades da mesma língua não é apenas política, mas ontológica. Critérios linguísticos de inteligibilidade mútua e sociolinguísticos de identidade comunitária frequentemente divergem. Português e galego são línguas distintas? Servo-croata era uma língua ou duas? Essas questões não têm resposta puramente empírica.

A observação de Max Weinreich (1945), “uma língua é um dialeto com um exército e uma marinha”, captura precisamente a dimensão política da individuação linguística. O que conta como língua distinta versus dialeto depende, em grande medida, de fatores extra-linguísticos como soberania política, padronização escolar e prestígio institucional. Isso tem implicação filosófica direta, dado que a ontologia das línguas não é puramente naturalística, mas parcialmente normativa e socialmente construída.

E linguistas que documentam línguas ameaçadas enfrentam alguns dilemas éticos sobre se devem intervir para revitalizar as línguas ou apenas as documentar. Eles têm obrigações para com as comunidades falantes ou apenas para com a ciência? Como balancear a preservação cultural contra a autodeterminação comunitária? Esses dilemas não se resolvem por métodos empíricos.

A reflexão filosófica pode e deve informar, questionar e aprimorar a pesquisa linguística. Quando linguistas entendem os fundamentos filosóficos de suas práticas, tornam-se mais capazes de justificar suas escolhas teóricas, antecipar objeções, criticar normas estabelecidas e, inclusive, integrar subdisciplinas.

Este capítulo examina três domínios filosóficos centrais para a linguística, a saber, metafísica, epistemologia e ética. Em cada domínio, examino os principais problemas filosóficos, avalio posições concorrentes, e proponho caminhos defensáveis. Ao final, retomo a conexão com a antropologia, que nos deu a cultura, e aponto para o próximo capítulo, sobre matemática, onde a linguagem se torna ainda mais abstrata, transformando-se em estruturas ideais que parecem habitar um reino próprio.

2. O que é uma Língua?

A questão metafísica fundamental para a linguística é o que é uma língua. Essa pergunta aparentemente simples revela-se surpreendentemente difícil e controversa. Diferentes respostas implicam ontologias radicalmente diferentes e têm consequências profundas para a pesquisa linguística.

Línguas são objetos abstratos, como os números e as proposições, existindo independentemente de instâncias físicas? Ou são padrões comportamentais, redutíveis a disposições de falantes para produzir e compreender enunciados?

Ou ainda entidades sociais, constituídas por práticas comunitárias e convenções? Cada resposta enfrenta suas dificuldades.

Uma posição é o *platonismo linguístico*. Nele, as línguas são objetos abstratos que existem independentemente dos falantes, do tempo e do espaço. Assim como o número 7 existe independentemente de haver sete objetos físicos, a língua portuguesa existiria independentemente de haver falantes de português. Nessa visão, falantes não criam a língua, mas a descobrem ou instanciam.

O platonismo linguístico tem um antecedente na concepção saussuriana de *langue* como sistema abstrato, distinto da *parole* (o uso concreto). Saussure (1916/2006) não era platonista no sentido técnico, mas a distinção *langue*/*parole* estrutura o mesmo problema, o de que a língua como objeto de estudo científico parece ser algo diferente de qualquer enunciado concreto, algo que persiste através dos usos individuais. A questão filosófica é se essa "persistência" implica existência autônoma e abstrata ou pode ser explicada em termos mais mundanos de práticas sociais e competências cognitivas.

Essa posição explica por que diferentes falantes podem falar "a mesma língua" apesar de diferenças de dialeto. Eles instanciam a mesma entidade abstrata. Dessa forma, justificamos por que podemos estudar línguas mortas, dado que elas continuam existindo como objetos abstratos, mesmo sem falantes, harmonizando-se com certas intuições da linguística gerativa, na qual gramáticas são sistemas formais abstratos.

O problema do platonismo é a sua extravagância ontológica de postular entidades além do necessário. Por que postular objetos abstratos misteriosos quando podemos explicar fenômenos linguísticos em termos de entidades menos controversas, como cérebros, comportamentos, ou convenções sociais? O platonismo tornaria misteriosa a relação entre a língua abstrata e seu uso concreto. Como falantes acessam objetos abstratos para produzir enunciados físicos? E, ainda, se as línguas são abstratas, quantas existem? Infinitas? Por que português e espanhol são línguas distintas ao invés de variedades de uma língua românica abstrata? O platonismo linguístico, embora filosoficamente interessante, parece incompatível com o naturalismo científico que caracteriza a linguística contemporânea. A maioria dos linguistas rejeita, implícita ou explicitamente, o compromisso com entidades abstratas platônicas.

No extremo oposto, encontramos o nominalismo behaviorista, defendido por Quine (1960), e Bloomfield (1933) antes dele. Nesta visão, não existem línguas como entidades, apenas falantes individuais com disposições para produzir e responder a estímulos verbais. Falar "a mesma língua" é apenas ter disposições comportamentais suficientemente similares. Não há nada além de organismos individuais e seus comportamentos.

Essa posição tem a vantagem da parcimônia ontológica extrema, pois não postula nada além de organismos físicos e seus comportamentos observáveis. É compatível com o naturalismo científico forte e evita compromissos com entidades abstratas ou mentais controversas. No entanto, ele também subestima sistematicamente a criatividade linguística, dado que falantes produzem e compreendem sentenças nunca antes encontradas, não por condicionamento a estímulos, mas por aplicação de regras generativas, e não está claro como tal teórico poderia explicar tal coisa. Para ele, erros de fala não refletem a língua do falante, apenas falhas de performance, sem que seja explicado por que os falantes de uma comunidade têm disposições tão similares.

O behaviorismo linguístico foi amplamente rejeitado após a revolução chomskyana, e poucos linguistas contemporâneos o defendem. Mas algumas de suas intuições centrais, como as do naturalismo, da parcimônia e do foco no comportamento observável, permanecem influentes.

A refutação do behaviorismo linguístico por Chomsky (1965, 1986), em sua resenha de "Verbal Behavior" de Skinner, constitui um dos episódios mais citados na história da ciência cognitiva. Chomsky (1965, 1986) mostrou que o comportamento verbal não pode ser descrito adequadamente em termos de estímulo-resposta-reforço sem que esses termos se tornem tão vagos que perdem qualquer conteúdo explicativo. A "pobreza do estímulo" (*poverty of the stimulus argument*), a observação de que crianças adquirem gramáticas muito mais ricas do que o *input* recebido, poderia ser justificada por indução, tornou-se o argumento central para o inatismo e permanece controverso até hoje.

Contudo, a posição dominante na linguística contemporânea, especialmente na tradição gerativa, é o psicologismo mentalista, que nos diz que as línguas são sistemas cognitivos internalizados em cérebros individuais. Uma língua é, essencialmente, uma gramática mental, i.e., um sistema de regras e representações implementado neurologicamente no cérebro do falante. Chomsky (1986)

distingue a língua-I, que seria interna, individual e intensional, da língua-E, que seria externa, social e extensional, argumentando que apenas a língua-I é objeto legítimo da ciência linguística.

Essa posição explica a criatividade linguística, pois nos diz que os falantes geram sentenças novas aplicando regras mentais. Dessa forma, nos explica a distinção competência-performance, indicando que a competência é o sistema mental subjacente, e a performance é o seu uso em tempo real. Os falantes de uma comunidade têm gramáticas mentais similares devido à dotação genética compartilhada e à exposição linguística similar. E é compatível com o naturalismo, ao sustentar que as gramáticas mentais são implementadas fisicamente em cérebros, sendo assim objetos naturalísticos investigáveis pelas ciências cognitivas.

Entretanto, tal posição torna problemática a individuação de línguas, pois se línguas são gramáticas mentais individuais, então cada falante tem sua própria língua ou linguagem privada. Por que e como dizer que dois falantes falam "a mesma língua"? Os gerativistas respondem que falantes têm gramáticas suficientemente similares para permitir a comunicação, mas "suficientemente similar" é algo extremamente vago e contextual. E as línguas mudam historicamente, têm prestígio diferencial, marcam identidades sociais, e nenhum desses fenômenos é explicável pela psicologia individual. Na verdade, nem o fato de que os linguistas conhecem a língua é explicado adequadamente pela teoria, pois não é claro como a língua é conhecida pelos linguistas, se as línguas são gramáticas mentais inconscientes. Ainda que a resposta gerativista apele à inferência abdutiva a partir dos comportamentos linguísticos, ela tornaria a língua menos diretamente acessível do que ela nos parece.

Uma alternativa ao psicologismo é o institucionalismo social. Tal teoria nos diria que línguas são instituições sociais, constituídas por práticas comunitárias, convenções, e normas sociais. Searle (1995) desenvolve essa abordagem no contexto da ontologia social. As línguas existem do mesmo modo que o dinheiro, as leis ou as universidades existem; eles não são objetos físicos, nem objetos mentais individuais, mas fatos institucionais criados por aceitação coletiva.

Nessa visão, dizer que "português é uma língua" é como dizer que "o real é moeda oficial do Brasil"; é um fato institucional, criado e mantido por práticas sociais de reconhecimento e uso. Línguas têm normas (gramaticais, fonológi-

cas, pragmáticas) análogas a normas legais ou monetárias; elas existem porque uma comunidade as reconhece e as trata como vinculantes, não porque estão inscritas em cérebros individuais ou em objetos abstratos platônicos.

A analogia entre línguas e instituições monetárias é fértil mas tem limites. Uma diferença crucial é que as normas monetárias são tipicamente explícitas e codificadas (leis, decretos), enquanto as normas gramaticais são, na sua maioria, implícitas e não articuladas pelos falantes. Isso levanta a questão de como normas implícitas podem constituir uma instituição no sentido de Searle (1995). Uma resposta possível é apelar para o conceito de "regras constitutivas" (*constitutive rules*) de Searle (1995), a saber, assim como as regras do xadrez constituem o jogo sem precisar ser conscientemente conhecidas por todos os jogadores, as normas gramaticais constituem a língua sem precisar ser explicitamente articuladas.

O institucionalismo explica naturalmente a dimensão social da linguagem, pois dá conta da variação sociolinguística, da mudança histórica, da distinção entre línguas de prestígio e línguas estigmatizadas, das políticas linguísticas e da identidade linguística; tudo isso faz sentido se as línguas são instituições sociais. E também faz sentido o fato de que a individuação das línguas é parcialmente arbitrária e politicamente carregada; a distinção entre "língua" e "dialeto" não seria puramente linguística, mas sociopolítica, assim como a distinção entre "moeda" e "vale refeição" não é puramente econômica.

Parece que o institucionalismo pode nos dar tudo que queríamos de uma teoria das línguas, porém note que há múltiplos desafios a serem vencidos pela teoria. Um deles é o seu aspecto ontologicamente inflacionário, por postular fatos institucionais como uma categoria metafísica adicional para além dos fatos físicos e mentais. E se línguas são convenções sociais, por que não encontramos línguas com qualquer estrutura imaginável? Os Institucionalistas poderiam até responder que as restrições vêm das capacidades cognitivas humanas, contudo, se fizerem isso, trazem de volta a dimensão psicológica como explanatoriamente ineliminável.

Um problema metafísico central aqui é a tensão entre universalidade e diversidade. Por um lado, todas as línguas humanas compartilham propriedades notáveis, como o fato de todas terem sintaxe recursiva, e de haver distinção nome-verbo, sistemas pronominais, mecanismos de referência. Por outro lado,

línguas diferem espetacularmente na ordem das palavras, na marcação morfológica, nos sistemas de sons, nas estratégias de construção de sentenças complexas. Como reconciliar universalidade e diversidade?

Chomsky (1965, 1986) propõe a Gramática Universal (GU), que seria um conjunto de princípios e parâmetros inatos que estruturam todas as línguas humanas. Os princípios são universais e invariáveis, como, por exemplo, nos casos da estrutura recursiva e das dependências sintáticas das línguas humanas. E os parâmetros são dimensões de variação que cada língua fixa para os seus próprios falantes, mas que podem variar entre línguas, como, por exemplo, a ordem núcleo-complemento, a existência de sujeito oculto etc. A GU seria parte da dotação genética humana, um módulo cognitivo especializado para a linguagem.

Essa hipótese explica vários fenômenos intrigantes, como o fato de as crianças adquirem gramáticas complexas a partir de *inputs* pobres e limitados, a partir de tomar muito da estrutura gramatical como inata. Assim, todas as línguas compartilhariam certos traços, porque todos os humanos compartilham a mesma GU. Contrariamente, certas estruturas logicamente possíveis nunca ocorrem em nenhuma língua porque violam os princípios da GU.

No entanto, a hipótese da GU é altamente controversa. Estudos de aquisição da linguagem mostram que o *input* linguístico é mais rico e sistemático do que os gerativistas assumem; e crianças são aprendizes estatísticos sofisticados que podem extrair padrões complexos de dados. Ademais, postular um módulo inato específico para a linguagem é biologicamente implausível, pois não há evidência clara de genes específicos para a gramática, e a evolução de um módulo tão específico é difícil de sustentar.

O desafio empírico mais contundente à GU chomskyana veio do estudo da língua Pirahã, conduzido por Daniel Everett (2005), que afirma que essa língua não tem recursividade sintática, o que violaria o princípio mais fundamental da GU. Chomsky (1965) e colaboradores contestam essa interpretação, argumentando que a ausência aparente de recursividade é performance, não competência. O debate ilustra como a controvérsia em torno da GU não é apenas empírica, mas metodológica e filosófica. O que contaria como evidência decisiva para ou contra a hipótese inatista?

Uma alternativa à GU seria o emergentismo funcionalista, defendido por Bybee (2010) e Tomasello (2003), entre outros. Nessa visão, os universais linguísticos não refletem uma gramática universal inata, mas emergem das restrições cognitivas gerais advindas da memória, da atenção e da categorização, e das restrições pragmáticas, como os processos de gramaticalização e mudança linguística. Por exemplo, todas as línguas distinguem nomes de verbos não porque essa distinção está *hardwired* em nós, mas porque ela reflete uma distinção cognitiva fundamental entre objetos e ações, e porque é funcionalmente útil para comunicação. A recursividade, nessa teoria, não seria inata, mas emergeria de processos gerais de combinação que operam em vários domínios cognitivos; e os parâmetros não seriam geneticamente codificados, mas pontos de variação que surgem de diferentes soluções para pressões funcionais similares.

O emergentismo evita o compromisso ontológico com módulos cognitivos inatos específicos para linguagem, sendo mais parcimonioso ontologicamente e compatível com a evidência de continuidade entre a linguagem humana e sistemas comunicativos de outras espécies. E explicaria a diversidade linguística naturalmente, ao dizer que diferentes línguas emergem de múltiplas pressões funcionais.

Todavia, o emergentismo não explica tão claramente a uniformidade e a rapidez da aquisição da linguagem, dado que, se as gramáticas emergem de processos gerais de aprendizagem, por que a aquisição é tão uniforme e rápida? Os emergentistas respondem normalmente que o *input* é mais estruturado e a aprendizagem mais poderosa do que se pensava; porém o problema dessa resposta é que ela é empírica e contestável, não explicando tão claramente por que certas estruturas logicamente possíveis nunca ocorrem, dado que há múltiplas pressões das quais as estruturas linguísticas emergem. Os emergentistas podem até apelar para os custos cognitivos ou funcionais, mas essas explicações são frequentemente *post hoc*.

Face às dificuldades de cada posição, proponho o que chamo de realismo linguístico modesto, uma posição intermediária, inspirada em debates filosóficos sobre o realismo científico e análoga ao realismo cultural modesto defendido no capítulo anterior. O realismo linguístico modesto sustenta que as línguas são reais, não meras ficções convenientes; mas que o modo de existência delas é

complexo e irredutível a uma única categoria ontológica. Inclusive, as próprias línguas têm aspectos multimodais ineliminâveis ou irredutíveis aos outros, como o aspecto psicológico da gramática mental, o aspecto social de fazer parte das práticas e normas comunitárias, e o aspecto abstrato de ser um sistema formal estudado por linguistas. De fato, são diversos sistemas formais, advindos de uma combinação de restrições cognitivas gerais, restrições funcionais da comunicação, e processos históricos de convergência; e não variações de uma única estrutura profunda universal, de modo que não é necessário postular um módulo inato altamente específico para as línguas ou para a competência linguística.

Essa posição rejeita o platonismo robusto que postula línguas como objetos abstratos independentes, mas admite que as descrições formais das línguas têm um estatuto abstrato similar aos modelos matemáticos em física. Não aceitamos nem o psicologismo puro, que reduz a língua à gramática mental individual, mas aceitamos que a competência linguística individual é psicologicamente real; e nem aceitamos o institucionalismo puro, que ignora os aspectos psicológicos, dado que tomamos a dimensão social da língua como ineliminável.

Como essa posição é justificada? Por analogia com o realismo científico. Elétrons são reais, mas o modo de existência deles é dado pela melhor teoria física, não por intuições ordinárias. Similarmente, línguas são reais, mas o modo de existência delas é dado pela melhor teoria linguística, não por intuições ordinárias sobre "línguas" como entidades unitárias simples. As melhores teorias linguísticas, sejam gerativas, funcionais ou sociolinguísticas, comprometem-se todas com múltiplas dimensões do linguístico, como o mental, o social e o formal. O realismo modesto respeita esse compromisso sem forçar reduções artificiais.

A justificação do realismo linguístico modesto por analogia com o realismo científico converge com a posição de Paul Postal (2003), que argumenta que objetos linguísticos como frases e sentenças são entidades abstratas, análogas a objetos matemáticos, mas causalmente eficazes no sentido de que estruturam o comportamento dos falantes. Essa posição é distinta do platonismo ingênuo, pois não exige que entidades linguísticas existam em um reino platônico independente, mas apenas que as descrições formais da linguística tenham um compromisso ontológico genuíno, não redutível à psicologia individual ou à sociologia das práticas.

Este realismo modesto conecta-se diretamente com o capítulo anterior. Se a cultura é real, múltipla, emergente e histórica, a linguagem, como produto e veículo da cultura, herda essas propriedades. A linguagem não é um sistema abstrato pairando acima da sociedade, nem um mero epifenômeno de processos cognitivos individuais. Ela é, simultaneamente, cognitiva e social, individual e coletiva, estruturada e histórica. O pluralismo ontológico que emerge aqui ecoa o pluralismo da antropologia e prepara o terreno para a matemática, onde a abstração se torna ainda mais radical.

3. Como Conhecemos as Línguas?

Questões epistemológicas são igualmente centrais para a linguística. Como os linguistas conhecem as línguas que estudam? Quais métodos são apropriados? Que tipo de evidência é relevante? Como as teorias linguísticas são justificadas? Um problema epistemológico fundamental é saber que tipo de evidência é relevante para teorias linguísticas. Diferentes abordagens linguísticas privilegiam diferentes tipos de evidência, e essa diferença não é meramente metodológica, mas reflete desacordos epistemológicos profundos.

Na tradição gerativa, a evidência central são os julgamentos de gramaticalidade, i.e., as intuições de falantes nativos sobre se as sentenças são gramaticais, agramaticais, ou marginais. Os linguistas constroem teorias formais que predizem esses julgamentos. A metodologia é semelhante à física teórica, a saber, construir modelos formais elegantes que capturam padrões nos dados, permitindo idealização, abstração e previsão.

Os julgamentos de gramaticalidade fornecem acesso direto à competência linguística, não apenas à performance, e permitem estudar fenômenos raros ou inexistentes em *corpora* naturais, mas ainda assim parte da competência linguística. Todavia, os julgamentos de gramaticalidade são frequentemente não confiáveis, dado que diferentes falantes julgam diferentemente e até o mesmo falante julga diferentemente em ocasiões diferentes. Além disso, os julgamentos refletem não apenas a competência sintática, mas também pragmática, semântica, estilística e até ideológica, privilegiando os linguistas como informantes, ainda que eles não sejam falantes típicos.

Alternativamente, as abordagens funcionalistas e baseadas no uso da linguagem privilegiam dados de *corpus*, i.e., amostras de uso linguístico real, falado ou escrito. A metodologia seria semelhante à biologia de campo porque os organismos são observados em ambientes naturais ao invés de laboratórios controlados. A vantagem aqui é que os dados de *corpus* refletem o uso real, e não intuições abstratas, calibrando vieses de julgamentos conscientes com dados empíricos do *corpus*.

Mas dados de *corpus* também têm limitações. O *corpus* nunca é exaustivo, de modo que a ausência de uma estrutura num *corpus* linguístico não prova que ele é agramatical, apenas que é raro. Assim, o corpus contém erros, disfluências, e estruturas de performance que não refletem a competência linguística; e torna-se difícil, então, usá-lo para avaliar a gramaticalidade de uma expressão ou a competência linguística. Dessa forma, ainda não está claro qual o tipo de evidência devemos aceitar ao avaliar as línguas.

Psicolinguistas, por sua vez, usam métodos experimentais, como medição de tempo de reação, rastreamento ocular, estudos de neuroimagem, entre outros. Esses métodos fornecem evidência sobre o processamento linguístico em tempo real, não apenas sobre o produto final ou os julgamentos de gramaticalidade. Como o tempo de reação é uma medida objetiva, isso torna a teoria não sujeita a vieses conscientes, com experimentos que permitem um controle rigoroso e sua replicabilidade. Entretanto, os experimentos acabam medindo a performance linguística, e não a competência diretamente, dado que o tempo de reação reflete algo sobre o tempo de processamento, mas a relação entre processamento e representação mental subjacente é indireta. E os experimentos têm uma desvantagem advinda de sua própria natureza experimental, que é não refletir uso linguístico natural.

Então, até agora, temos pelo menos três teorias sobre a natureza das evidências em linguística, ao nos perguntarmos sobre como adquirimos conhecimento em linguística. Todas elas têm desvantagens e vantagens, e não são compatíveis entre si. Temos uma dificuldade real de escolher entre elas. Um problema epistemológico que se coloca aqui é o problema da subdeterminação empírica (ver Quine, 1960; Duhem, 1906/1991), dado que teorias linguísticas incompatíveis podem ser igualmente compatíveis com toda evidência empírica possível. Como escolher entre teorias empiricamente equivalentes?

Esse problema não é exclusivo da linguística, mas ele se manifesta de uma forma particularmente aguda nela. Vamos considerar um exemplo para sentir a dificuldade na prática. Pense em duas teorias que tentam explicar como construímos sentenças. De um lado, temos as gramáticas transformacionais gerativas, a tradição iniciada por Chomsky (1965). Elas postulam que a estrutura de uma sentença tem diferentes níveis (uma estrutura profunda e uma estrutura superficial) e que regras de transformação conectam um nível ao outro. De outro lado, temos as gramáticas de estrutura sintagmática generalizada (GPSG), que tentam fazer a mesma coisa, mas sem recorrer a transformações. Elas usam um sistema de regras muito mais complexo e interligado para gerar as sentenças diretamente. Ambas as teorias podem ser capazes de gerar exatamente o mesmo conjunto de sentenças para uma dada língua, ou seja, são empiricamente equivalentes. Elas descrevem os mesmos dados. No entanto, os mecanismos que postulam são radicalmente diferentes. Uma diz que a mente tem estruturas profundas ocultas e regras que as movimentam. A outra diz que não, que a estrutura é gerada de uma vez só por um sistema de regras mais intrincado.

O mesmo tipo de dilema aparece quando comparamos famílias teóricas ainda mais distantes. Considere as teorias gerativas (como a de Chomsky, 1965) e as teorias funcionais da ordem das palavras. Ambas podem, em muitos casos, fazer as mesmas predições sobre a ordem dos termos em uma frase. Mas as explicações que oferecem para por que essa ordem é assim são de naturezas completamente diferentes. A teoria gerativa dirá que é por causa de um princípio inato da Gramática Universal, uma restrição sintática que vem “de fábrica” no cérebro humano. A teoria funcional dirá que é porque essa ordem específica atende melhor a pressões comunicativas, como colocar a informação mais importante no início da frase para facilitar a compreensão do ouvinte. Temos, novamente, a mesma capacidade preditiva, mas explicações causais que apontam para direções opostas, uma para dentro (a mente inata) e outra para fora (a função social da comunicação).

Diante dessa situação, que os filósofos da ciência chamam de subdeterminação da teoria pela evidência, quando os dados não são suficientes para determinar qual teoria é a verdadeira, como podemos fazer uma escolha racional? Os filósofos da ciência sugerem que, quando a evidência empírica não decide,

somos forçados a apelar para o que chamam de virtudes teóricas. São como critérios de qualidade que uma boa teoria deve ter.

A simplicidade faz com que a teoria mais simples, que postula menos (tipos de) entidades ou regras, seja geralmente preferível; é a navalha de Ockham, i.e., não multiplicar os seres sem necessidade. Pelo poder explanatório, uma teoria que explica uma ampla gama de fenômenos é melhor do que uma que explica apenas um punhado deles. A fertilidade, por sua vez, é a virtude de uma boa teoria que não apenas explica o que já sabemos, mas também sugere novas perguntas, abre novos caminhos de pesquisa e é "frutífera". Pelo conservadorismo, é preferível uma teoria que se encaixa bem com o conhecimento que já temos em outras áreas, em vez de uma que force uma revisão radical de tudo o que sabemos. E ainda há a elegância matemática ou formal, pois uma teoria com uma estrutura matemática bela e coerente é, para muitos cientistas e filósofos, um sinal de que ela está no caminho certo.

A sistematização das virtudes teóricas como critérios de escolha entre teorias empiricamente equivalentes foi desenvolvida por Kuhn (1977) em *Objectivity, Value Judgment, and Theory Choice* e por Quine e Ullian (1970) em *The Web of Belief*. Ambos reconhecem que esses critérios são genuinamente objetivos, não são meras preferências subjetivas, mas também que sua aplicação requer julgamento e não é algorítmica. A linguística encontra-se exatamente nessa situação; nela, os critérios existem e são racionalmente defensáveis, mas não produzem decisões automáticas.

Os linguistas, na prática do dia a dia, usam essas virtudes o tempo todo, mesmo que implicitamente. Um gerativista acha sua teoria melhor porque ela é mais elegante e tem um grande poder unificador. Um funcionalista acha a sua melhor porque é mais simples e se conecta bem com a psicologia e a sociologia. O problema, e aqui entra a contribuição da filosofia, é que esses critérios raramente são discutidos de forma explícita ou justificada. Um linguista pode preferir uma teoria por sua simplicidade, mas o que exatamente ele entende por simplicidade? Por que a simplicidade deveria ser um guia para a verdade?

Minha proposta é que a escolha teórica em linguística, longe de ser um processo oculto e subjetivo, deve ser guiada por uma reflexão explícita sobre essas virtudes. Ou seja, precisamos tornar os critérios visíveis e discuti-los abertamente. A simplicidade ontológica, por exemplo, que significa postular o menor núme-

ro possível de entidades (tipos de regras, níveis de representação), é um valor importante. No entanto, ele não pode ser um valor absoluto. Uma teoria que seja um pouco mais complexa, mas que tenha um poder explanatório significativamente maior (que explique fenômenos que a teoria simples deixa de fora), pode ser perfeitamente justificável. A unificação explanatória, que é a capacidade de explicar fenômenos muito diversos (como a ordem das palavras em diferentes línguas) com um punhado de princípios comuns, é também uma virtude poderosa. Mas ela não pode ser buscada a ponto de distorcermos os dados para que eles se encaixem na nossa teoria. A fertilidade é crucial, pois uma ciência que não gera novas perguntas e não estimula novas pesquisas é uma ciência que está morrendo. E, claro, a adequação empírica continua sendo a base de tudo, pois se uma teoria não consegue sequer descrever os dados básicos, as outras virtudes não importam.

O ponto crucial, e que exige uma boa dose de maturidade intelectual, é que essas virtudes frequentemente entram em conflito. Uma teoria pode ser mais simples, mas menos fértil. Outra pode ter um enorme poder explanatório, mas ser extremamente complexa. Não existe um algoritmo, uma fórmula mágica, que nos diga como pesar esses valores e chegar a uma resposta definitiva. A escolha teórica, portanto, envolve um julgamento informado, uma capacidade de ponderar prós e contras que é mais próxima da decisão de um juiz em um caso complexo do que da aplicação de uma regra matemática. Isso pode soar desconfortável para quem espera da ciência uma precisão absoluta, mas não torna a linguística uma atividade subjetiva ou arbitrária. A linguística, assim como as humanidades, tem uma dimensão interpretativa ineliminável. Isso não a torna menos científica; apenas torna a sua racionalidade mais rica e complexa do que a mera dedução lógica.

Outra distinção epistemológica fundamental, e que nos ajuda a clarear o terreno, é a que separa dois tipos radicalmente diferentes de conhecimento envolvidos na linguística. De um lado, temos o conhecimento do falante nativo. Um falante de português "conhece" sua língua no sentido mais profundo, dado que ele é capaz de produzir e compreender um número infinito de sentenças que nunca ouviu antes. Ele consegue julgar, intuitivamente, se uma frase soa "estranha" ou "agramatical" ("eu a vi" comparado com "a eu vi"). Ele detecta erros. Mas se você perguntar a ele por que "a eu vi" é errado, ou qual é a regra que

ele está usando, ele provavelmente não saberá explicar. Esse conhecimento é, em sua maior parte, inconsciente, implícito, não articulável, tácito. É uma habilidade prática.

Por outro lado, temos o conhecimento do linguista teórico. O linguista pode não ser falante nativo do português, mas ele é capaz de descrever e explicar as estruturas da língua de forma explícita. Ele pode formular a regra que coloca os pronomes oblíquos antes do verbo em certos contextos e depois em outros. Ele pode construir uma teoria gramatical completa e coerente. Esse conhecimento é consciente, explícito, proposicional. É um "saber que" as coisas são de tal e tal jeito.

Essa distinção é crucial e foi brilhantemente articulada pelo filósofo Gilbert Ryle (1949), que a cunhou como a diferença entre "*knowing how*" (saber como) e "*knowing that*" (saber que). Saber andar de bicicleta é um "*know-how*", é uma habilidade corporal, um conhecimento prático que não depende de ser capaz de articular as leis da física envolvidas no equilíbrio. Saber que a capital da França é Paris é um "*know-that*", é um conhecimento proposicional, um fato que podemos declarar. O conhecimento do falante é um "*know-how*" primário. O conhecimento do linguista é um "*know-that*". A relação entre eles, no entanto, é o cerne de um grande debate.

Para Chomsky (1986), essa relação é direta e constitutiva. Ele argumenta que as teorias linguísticas que construímos (o "*know-that*" do linguista) são, na verdade, descrições formalizadas do conhecimento implícito do falante (o seu "*know-how*"). A gramática gerativa que o linguista escreve no papel seria, segundo essa visão, uma teoria da "gramática mental" que está no cérebro do falante. O conhecimento do falante e do linguista teriam, portanto, o mesmo objeto (a gramática mental), mas modos de acesso diferentes (um implícito e inconsciente, outro explícito e consciente). O linguista estaria, em última análise, "traduzindo" o que o falante sabe fazer para uma linguagem teórica.

Críticos dessa posição levantam uma objeção poderosa. Eles argumentam que Chomsky (1986) está confundindo duas coisas muito diferentes. A capacidade de fazer algo não implica a existência de um conhecimento proposicional inconsciente sobre como se faz. Eu posso saber andar de bicicleta perfeitamente sem que, em algum lugar da minha mente, exista um tratado de física e fisiologia esperando para ser descoberto. O "saber como" é uma capacidade, uma

habilidade, não um conhecimento proposicional oculto. Da mesma forma, eu sei falar português, mas isso não significa que eu tenha, em algum nível inconsciente, uma teoria sintática completa e coerente armazenada no meu cérebro. As teorias linguísticas, para esses críticos, não descrevem um conhecimento implícito; elas constroem modelos abstratos que são úteis para simular e explicar a competência linguística, da mesma forma que um mapa não é o território, mas uma representação útil dele.

Diante desse impasse, proponho uma posição intermediária, que me parece mais acurada e menos problemática. Podemos pensar que as teorias linguísticas descrevem estruturas que são funcionalmente realizadas nas competências linguísticas dos falantes. Em outras palavras, o cérebro/mente do falante implementa um sistema que se comporta como se seguisse as regras da teoria, mas isso não significa que o falante "conheça" essas regras em qualquer sentido, nem consciente nem inconscientemente.

Uma analogia pode ajudar. As teorias da biologia molecular descrevem os processos metabólicos que ocorrem dentro de uma célula. Sabemos que as enzimas interagem, que o DNA é transcrito em RNA, que a glicose é quebrada para gerar energia. Mas seria estranho dizer que a célula "conhece" essas teorias bioquímicas. A célula simplesmente realiza esses processos. O metabolismo é uma propriedade do sistema, não um conhecimento que o sistema possui. Da mesma forma, as teorias sintáticas descrevem os padrões de funcionamento da nossa competência linguística. Nós, como falantes, realizamos esses padrões. Mas não "conhecemos" as teorias que os descrevem. O conhecimento linguístico, nesse sentido, parece algo *sui generis*. Não se encaixa perfeitamente na dicotomia "saber como" / "saber que". Talvez possamos chamá-lo de "competência estruturada", i.e., uma capacidade complexa e altamente organizada, mas que não é, em si mesma, um corpo de conhecimento proposicional acessível à consciência ou à introspecção.

Se aceitarmos essa visão, as implicações práticas são imediatas. Se múltiplos tipos de evidência são legítimos e nos dão acesso a diferentes aspectos desse fenômeno complexo que é a linguagem, então a linguística precisa, necessariamente, ser uma disciplina metodologicamente pluralista. Julgamentos de gramaticalidade nos dão pistas sobre a competência idealizada. Dados de *corpus* nos mostram a língua em seu uso real, cheio de variações e nuances.

Experimentos psicolinguísticos revelam os processos em tempo real que acontecem quando falamos e ouvimos. Essas fontes de evidência não são rivais que disputam a primazia. Pelo contrário, elas são complementares, como diferentes instrumentos musicais em uma orquestra. Cada uma ilumina um aspecto diferente do fenômeno linguístico. O grande desafio para o linguista, e aí reside a arte da pesquisa, é saber integrar esses diferentes tipos de evidência em teorias coerentes, em vez de simplesmente escolher um tipo em detrimento dos outros.

Essa conclusão, como não poderia deixar de ser, ecoa fortemente o pluralismo metodológico que defendemos de forma tão detida no capítulo anterior, dedicado à antropologia. Lá, vimos que a antropologia, para fazer justiça à complexidade de seu objeto, a cultura, precisa lançar mão de métodos interpretativos para desvendar teias de significado, de métodos causais para explicar processos materiais e de adaptação, e de métodos evolutivos para compreender as transformações em escalas de tempo longas. Aqui, na linguística, encontramos um quadro rigorosamente análogo. Precisamos de métodos baseados na introspecção e no julgamento do falante (como na tradição gerativa), de métodos de observação naturalística e coleta de corpus (como na sociolinguística e na linguística funcional) e de métodos de experimentação controlada (como na psicolinguística). Nenhum desses métodos, isoladamente, é suficiente para capturar a totalidade e a riqueza do fenômeno linguístico. A linguística, assim como a antropologia, é, em sua própria essência, uma disciplina intrinsecamente pluralista. E é precisamente essa diversidade de abordagens e perspectivas que, longe de representar uma fraqueza ou um sinal de imaturidade científica, constitui a sua maior força e a fonte mais rica de seu contínuo desenvolvimento.

4. Justiça Linguística e Responsabilidades Profissionais

Quando discutimos filosofia da linguística, é muito comum que as questões éticas fiquem em segundo plano, como se fossem um adendo opcional à "verdadeira" investigação científica sobre a linguagem. Mas essa atitude, me parece, é um erro. As questões éticas são tão fundamentais quanto as ontológicas e as epistemológicas. Afinal, a linguística não é uma atividade que ocorre no vácuo. Ela é praticada por pessoas, em sociedades, e tem consequências diretas e profundas na vida de outras pessoas. Precisamos, portanto, nos perguntar se

os linguistas têm obrigações éticas específicas. Se sim, para com quem? E que valores devem orientar a sua prática de pesquisa, de ensino e de intervenção no mundo?

Um dos problemas éticos mais centrais, e que nos acompanhará ao longo de toda esta seção, é o da justiça linguística. O que isso significa? Basicamente, trata-se da ideia de que todas as línguas e todas as variedades de uma língua (os dialetos) merecem um tratamento justo, assim como os seus falantes. Isso pode parecer óbvio quando afirmado dessa forma, mas se choca frontalmente com as ideologias linguísticas que dominam a nossa sociedade e a maioria das sociedades do mundo. Essas ideologias são sistemas de crenças que hierarquizam as línguas e os modos de falar. Elas rotulam alguns dialetos como "corretos", "cultos", "superiores" ou "lógicos", enquanto condenam outros como "incorretos", "vulgares", "inferiores", "preguiçosos" ou "sem gramática". Pense no prestígio associado ao chamado "português padrão" em contraste com o estigma muitas vezes dirigido às variedades populares, rurais ou periféricas, como o chamado "dialeto caipira" ou as variações linguísticas das comunidades urbanas marginalizadas.

Ora, a posição da linguística científica, desde pelo menos o início do século XX com os trabalhos de Ferdinand de Saussure (1916/2006) e o desenvolvimento da tradição descritivista, é a de rejeitar veementemente essas hierarquias de valor. Para o linguista, não existe língua ou dialeto "melhor" do que outro do ponto de vista estritamente linguístico. Todos os dialetos de todas as línguas são sistemas igualmente complexos, igualmente estruturados, igualmente governados por regras (ainda que regras diferentes), igualmente expressivos e igualmente adequados para as necessidades cognitivas e comunicativas de seus falantes. As diferenças que percebemos entre uma "língua padrão" e um "dialeto não padrão" não são diferenças linguísticas inerentes, mas sim diferenças sociais. A variedade padrão é simplesmente aquela que, por contingências históricas e políticas, passou a ser falada pelos grupos socialmente dominantes, foi institucionalizada pelo sistema de ensino, pela mídia e pelo Estado, e adquiriu, por isso, um prestígio artificial. Chamar um dialeto de "incorreto" ou "pobre" é um ato tão desprovido de base científica quanto seria chamar uma espécie

biológica de “feia” ou “imperfeita”. É um juízo de valor travestido de constatação objetiva[1].

No entanto, essa posição descritivista, por mais correta e fundamental que seja do ponto de vista científico, acaba gerando tensões práticas bastante incômodas. Porque, no mundo real, fora dos muros da academia, os falantes de dialetos não padrão enfrentam uma discriminação real e concreta. Eles são, com frequência, excluídos de oportunidades educacionais e profissionais. São estereotipados como menos inteligentes, menos competentes ou menos preparados. Um jovem que chega a uma entrevista de emprego falando seu dialeto nativo pode ser julgado negativamente, mesmo que tenha todas as qualificações técnicas para a vaga. Diante dessa realidade brutal, o linguista que se limita a repetir, de forma um tanto abstrata, que “todos os dialetos são iguais”, pode acabar parecendo alguém que está ignorando ou minimizando o sofrimento real dessas pessoas. A afirmação científica, por si só, não elimina o preconceito e a discriminação.

Para complicar ainda mais o quadro, há o fato de que os próprios pais e membros de comunidades linguisticamente marginalizadas, muitas vezes, desejam ardentemente que seus filhos aprendam e dominem a variedade padrão. Eles percebem, com toda a razão, que o domínio da língua de prestígio é um instrumento valioso, uma ferramenta de ascensão social e de acesso a direitos. Quando um linguista, movido pelas melhores intenções, desencoraja esse aprendizado em nome da preservação da autenticidade linguística, ele corre o sério risco de soar paternalista e de estar, na prática, contribuindo para a manutenção da desigualdade, ao negar às pessoas as ferramentas para navegar no mundo social tal como ele é, e não como gostaríamos que fosse.

---

[1] Aqui me lembro do termo “regência”, da gramática do Português. Regência é o nome que damos à regra que determina qual preposição deve acompanhar um verbo ou um nome, como quando escolhemos se dizemos “assistir ao filme” ou “assistir o filme”, “obedecer aos pais” ou “obedecer os pais”, “ter medo de altura” ou “ter medo por altura”. A etimologia da palavra “regência” é reveladora. Vem do latim *regere* (governar, conduzir), a mesma raiz de “rei” e “regente”. Isso não é acidental, pois a regência gramatical era, historicamente, a forma como o rei falava. A regência, esse tópico que atormenta estudantes, carrega, sem que percebamos, um vestígio da estrutura de poder que a moldou.

A tensão entre o descritivismo linguístico e as demandas práticas das comunidades marginalizadas foi tematizada por Hinton & Hale (2001) no contexto das políticas de revitalização, dado que reconhecer a igualdade linguística de todas as variedades não implica que todas as variedades tenham o mesmo valor instrumental no mercado linguístico. O conceito de "capital linguístico" de Bourdieu (1982) nos diz que as línguas e as variedades funcionam como capital simbólico, e a distribuição desigual desse capital é uma dimensão da reprodução das desigualdades sociais que o descritivismo linguístico, por si só, não resolve.

Perceba, então, que estamos diante de um dilema ético genuíno e de difícil resolução. De um lado, há um imperativo científico e moral de combater as ideologias linguísticas que perpetuam a injustiça, afirmando a igualdade fundamental de todas as variedades. De outro lado, há a necessidade de respeitar a agência e as escolhas das comunidades linguísticas, reconhecendo as realidades sociolinguísticas do mundo em que vivemos, incluindo o valor instrumental inegável que as variedades padrão possuem. Como podemos reconciliar esses dois compromissos, aparentemente conflitantes?

Uma maneira produtiva de enfrentar esse dilema, proponho, é adotar uma abordagem de justiça linguística que distingue duas dimensões complementares, inspiradas em debates da filosofia política, a dimensão do reconhecimento e a dimensão da redistribuição. O reconhecimento diz respeito à esfera simbólica e cultural. É a luta para que todas as variedades linguísticas sejam valorizadas, respeitadas e reconhecidas como expressões legítimas da identidade e da cultura de seus falantes. É o esforço para desconstruir o estigma e afirmar publicamente que uma fala "caipira" ou uma fala "periférica" não é inferior, é apenas diferente. A redistribuição, por sua vez, diz respeito à esfera material e econômica. É a luta para garantir que todos tenham acesso efetivo aos recursos linguísticos que conferem poder social e abrem portas, e isso inclui, sim, o acesso ao domínio pleno da variedade padrão. Uma concepção plena de justiça linguística, sustento, requer que atuemos em ambas as frentes simultaneamente.

Traduzindo isso em termos mais concretos, essa abordagem implica um compromisso ativo de combater as ideologias estigmatizantes. Isso significa educar o público, por meio de artigos, livros, entrevistas, participação em deba-

tes, sobre a natureza da linguagem e a igualdade fundamental entre as variedades linguísticas. Significa mostrar que o preconceito linguístico é uma forma de preconceito como qualquer outro, e que precisa ser combatido. Significa também garantir o acesso à variedade padrão, não como uma forma de erradicar o dialeto nativo do falante, mas como uma forma de adição bidialetal. O objetivo não é fazer com que uma criança abandone a língua de sua comunidade para aprender a "falar certo". O objetivo é que ela se torne capaz de transitar entre diferentes registros, de dominar tanto a sua variedade nativa quanto a variedade padrão, podendo usar uma ou outra conforme o contexto, ampliando assim seu repertório comunicativo e suas possibilidades de atuação social. Além disso, precisamos reformar as instituições para que se tornem mais inclusivas linguisticamente. Isso significa, por exemplo, repensar os manuais didáticos, a formação de professores, os critérios de avaliação, para que a variedade padrão não seja a única forma de expressão aceita e valorizada. Significa criar espaços onde a diversidade linguística seja não apenas tolerada, mas celebrada. Por fim, e talvez o mais importante, é preciso capacitar as comunidades para que elas próprias possam fazer suas escolhas. Em vez de o linguista, de fora, decidir o que é melhor para a comunidade (se deve preservar sua língua, se deve adotar a variedade padrão etc.), a função do pesquisador deve ser a de fornecer informações, compartilhar conhecimento e apoiar os processos de decisão que partam da própria comunidade, respeitando sua autonomia e sua capacidade de autodeterminação.

Essa tensão se torna ainda mais aguda quando consideramos um fato alarmante. Estima-se que, das cerca de sete mil línguas faladas atualmente no mundo, aproximadamente a metade esteja ameaçada de extinção ainda neste século (ver Nettle & Romaine, 2000; Evans, 2010). Diante desse cenário, os linguistas se veem confrontados por questões éticas de primeira grandeza. Eles têm a obrigação de documentar essas línguas ameaçadas? Devem intervir ativamente para tentar revitalizá-las, evitando seu desaparecimento? Ou, ao contrário, devem respeitar os processos de mudança linguística como fenômenos naturais, mesmo que isso signifique, em alguns casos, a extinção de uma língua?

De um lado, há um argumento forte em favor da documentação. Defensores dessa posição sustentam que as línguas são um patrimônio cultural inestimável

da humanidade. Cada língua é um repositório único de conhecimento, uma forma particular de ver, classificar e significar o mundo. Sua perda é uma perda irreparável para a diversidade cultural e para o conhecimento humano como um todo. Os linguistas, por possuírem as ferramentas e a expertise necessárias para registrar e descrever essas línguas, teriam, portanto, uma responsabilidade especial de agir antes que seja tarde demais, de criar um registro que preserve, para a posteridade, esse tesouro que está se perdendo. Além disso, essa documentação pode, em muitos casos, beneficiar as próprias comunidades falantes. Um dicionário bem feito, uma gramática descritiva, uma coletânea de narrativas tradicionais podem se tornar materiais valiosos para projetos de revitalização linguística no futuro, ou mesmo para o fortalecimento da autoestima e da identidade cultural da comunidade no presente.

Do outro lado, no entanto, há argumentos céticos igualmente respeitáveis. Críticos dessa posição "preservacionista" argumentam que a perda linguística é, em muitos aspectos, um processo natural de mudança cultural. As línguas sempre nasceram, se transformaram e morreram ao longo da história humana. Tentar preservar artificialmente uma língua que as próprias comunidades estão abandonando pode ser uma forma de paternalismo bem-intencionado, mas paternalismo ainda assim. Pode significar ignorar os desejos e as necessidades reais dessas comunidades, que podem, legitimamente, preferir adotar uma língua mais dominante como forma de acessar melhores oportunidades econômicas, educacionais e de participação social. Além disso, há o risco de que a documentação linguística se torne uma forma de exploração. O linguista, muitas vezes vindo de uma universidade rica, chega à comunidade, extrai o conhecimento linguístico (que é um bem imaterial da comunidade), publica artigos e livros em sua língua acadêmica, ganha prestígio e financiamento, e deixa para trás muito pouco, ou nada, em termos de retorno concreto para as pessoas que tornaram sua pesquisa possível.

Perceba, mais uma vez, o dilema genuíno. A extinção de uma língua é, sem dúvida, uma perda. Mas a mudança linguística, em si mesma, não é intrinsecamente ruim. E, acima de tudo, as comunidades falantes têm o direito fundamental à autodeterminação, inclusive linguística. Como o linguista pode navegar por esse campo minado de tensões éticas?

Proponho que a saída está em adotar uma ética de documentação linguística baseada na colaboração comunitária, no consentimento informado e no benefício mútuo. Em termos práticos, isso significa que a documentação só deve ser iniciada após a obtenção do consentimento livre e esclarecido das comunidades envolvidas. A comunidade precisa compreender claramente quais são os objetivos do projeto, quais métodos serão utilizados, e quais os usos potenciais dos materiais que serão produzidos. As comunidades não podem ser tratadas como meros "objetos de estudo" ou "informantes". Elas devem ser parceiras ativas em todas as etapas do processo, desde a definição do que será documentado até a forma como os resultados serão apresentados e divulgados. Os materiais resultantes da documentação (gravações, transcrições, análises) devem ser disponibilizados de forma acessível e útil para a própria comunidade, e não apenas depositados em arquivos acadêmicos inacessíveis. Sempre que possível, o projeto deve incluir a capacitação de membros da comunidade como documentadores, para que eles próprios possam dar continuidade ao trabalho de registro e valorização de sua língua. Se a comunidade manifestar o desejo de empreender esforços de revitalização linguística, o linguista deve estar pronto para apoiar esse processo com sua expertise, oferecendo ferramentas e orientação. No entanto, as decisões sobre os objetivos e métodos da revitalização devem ser tomadas pela comunidade, não impostas de fora. E, em última instância, se a comunidade, após ser devidamente informada, decidir que não deseja que sua língua seja documentada ou que não quer empreender projetos de revitalização, o linguista deve respeitar essa decisão, por mais que, pessoalmente, possa lamentar a perda de diversidade linguística.

Essa abordagem reconhece, de forma realista, que a documentação linguística não é, e nunca foi, uma atividade moralmente neutra. Ela é uma prática que envolve relações de poder, interesses divergentes e consequências profundas para as pessoas envolvidas. A função do linguista, nesse contexto, não é a de um técnico imparcial, mas a de um agente ético que precisa equilibrar valores tão importantes quanto a preservação da diversidade cultural, o respeito pela autodeterminação das comunidades e a responsabilidade inerente ao seu próprio conhecimento especializado.

Para além dessas questões específicas sobre línguas ameaçadas, há uma questão ética mais ampla que precisa ser enfrentada. Os linguistas têm a obri-

gação de promover a justiça social para além dos limites estritos de suas atividades acadêmicas? A sua expertise deve ser colocada a serviço de causas políticas e sociais mais amplas?

Há quem defenda que sim. Essa posição é particularmente forte em tradições como a sociolinguística crítica e a análise crítica do discurso. Para esses pesquisadores, a linguagem não é um objeto neutro e transparente. Ela é, acima de tudo, um instrumento de poder. As estruturas linguísticas, as ideologias que circulam sobre as línguas e as políticas linguísticas adotadas pelos Estados são, com frequência, mecanismos que perpetuam desigualdades sociais baseadas em raça, classe, gênero e outras categorias. Nessa perspectiva, os linguistas, justamente por possuírem um conhecimento aprofundado sobre o funcionamento da linguagem, teriam uma responsabilidade especial de usar esse conhecimento para identificar, denunciar e combater os usos opressivos da palavra. Isso poderia incluir desde a elaboração de laudos periciais em processos judiciais envolvendo discriminação linguística, até a participação em debates públicos sobre políticas educacionais e a produção de material didático que promova uma visão mais inclusiva e menos preconceituosa da diversidade linguística.

Do outro lado, há quem veja essa posição com grande ceticismo. Para esses críticos, a ciência, para merecer esse nome, deve ser objetiva e politicamente neutra. Misturar a investigação linguística com ativismo social, por mais nobres que sejam as intenções, seria um erro grave, que comprometeria a integridade e a credibilidade do conhecimento produzido. Os linguistas são especialistas em estrutura linguística, não em justiça social ou em política. A sua autoridade, que deriva de seu conhecimento técnico, não se estende a essas outras esferas, nas quais eles seriam apenas mais um cidadão com uma opinião, e não um especialista com uma verdade a ser imposta.

Essa tensão, como se pode perceber, reflete um debate muito mais amplo na filosofia da ciência sobre a possibilidade e os limites da neutralidade científica e sobre a responsabilidade social dos cientistas. Diante disso, qual seria a posição mais defensável?

Sugiro que uma posição equilibrada deve reconhecer tanto a possibilidade quanto as limitações de uma abordagem neutra. Por um lado, é fundamental afirmar que as descrições linguísticas podem e devem aspirar a um ideal de

objetividade. Descrever a estrutura fonológica de um dialeto, analisar as regras morfológicas de uma língua ou documentar um processo de mudança linguística não exige, em si mesmo, nenhum comprometimento com valores políticos particulares. O linguista, nessa tarefa descritiva, não deve distorcer os fatos para que eles se alinhem com suas preferências ideológicas.

Por outro lado, é preciso reconhecer honestamente que muitas das perguntas que a linguística se coloca têm dimensões normativas que são inescapáveis. A decisão sobre quais línguas documentar (e quais deixar de lado), a forma como classificamos determinados dialetos (se como línguas independentes ou como variedades de uma mesma língua), a escolha de qual variedade ensinar nas escolas como padrão, todas essas questões envolvem julgamentos de valor e não podem ser respondidas apenas com base em descrições empíricas. Além disso, o linguista não pode simplesmente ignorar as implicações sociais de sua própria pesquisa. Os resultados de seus trabalhos, uma vez publicados, passam a circular na sociedade e podem influenciar políticas educacionais, reforçar ou combater percepções públicas sobre dialetos, e afetar, para o bem ou para o mal, o tratamento que as comunidades linguísticas recebem.

Diante disso, acredito que os linguistas têm uma responsabilidade especial de considerar cuidadosamente as implicações sociais de seu trabalho. Isso não significa que devam se transformar em ativistas partidários ou que sua produção acadêmica deva ser pautada por uma agenda política pré-definida. Significa cultivar uma atitude de reflexividade, i.e., a capacidade de refletir criticamente sobre as dimensões normativas que estão sempre presentes, ainda que de forma implícita, na prática linguística, comunicando suas descobertas de uma maneira que não perpetue, ainda que involuntariamente, preconceitos e estereótipos e, quando for pertinente e relevante, contribuir para os debates públicos sobre políticas linguísticas, oferecendo não opiniões pessoais, mas o conhecimento especializado que domina.

5. Pluralismo Metodológico como Resposta

Ao longo deste capítulo, examinamos três ordens de problemas filosóficos que são absolutamente centrais para a linguística. Em cada um deles, o padrão se repetiu. Vimos que as posições extremas, por mais atraentes que possam

parecer à primeira vista, revelam-se insustentáveis quando submetidas a um exame mais cuidadoso. Sobre a natureza da língua, nem o platonismo que a trata como um objeto abstrato e eterno, nem o behaviorismo que a reduz a padrões de comportamento observável, nem o psicologismo puro que a confina à mente individual, nem o institucionalismo puro que a dissolve em convenções sociais são, isoladamente, capazes de dar conta de toda a complexidade do fenômeno. Diante disso, propus o que chamei de realismo linguístico modesto, uma posição que reconhece as múltiplas dimensões (cognitiva, social, formal) do linguístico como igualmente reais e irredutíveis umas às outras.

Em seguida, ao investigarmos as bases do conhecimento linguístico, vimos que os diferentes tipos de evidência disponíveis (os julgamentos de gramaticalidade, os dados de corpus, os experimentos psicolinguísticos) possuem cada um as suas vantagens e as suas limitações, e que nenhum deles pode, sozinho, reivindicar o status de fonte única e privilegiada de acesso à verdade sobre a língua. Essa constatação nos levou a defender um pluralismo metodológico que integre essas múltiplas fontes de evidência, reconhecendo sua natureza complementar em vez de rival.

Por fim, ao nos debruçarmos sobre as responsabilidades éticas dos linguistas, nos deparamos com dilemas que não admitem soluções simples. Tanto no debate sobre justiça linguística quanto na questão da documentação de línguas ameaçadas e no papel social mais amplo do pesquisador, vimos que valores legítimos entram em conflito e que é preciso fazer escolhas difíceis. A saída que propusemos passou pela adoção de uma ética baseada na colaboração, no respeito à autonomia das comunidades e na reflexividade sobre as implicações sociais da prática científica.

Se olharmos com atenção para o padrão que se repete nessas respostas, veremos um traço comum, a saber, a rejeição sistemática das posições extremas e a defesa de posições intermediárias que buscam preservar as intuições mais sólidas de cada um dos lados em conflito. Para alguns, isso pode parecer uma postura insatisfatória, como uma mera procura de síntese hegeliana, ou talvez uma espécie de ecletismo preguiçoso que tenta agradar a todos sem se comprometer com nada. No entanto, acredito que essa aparência engana. Essa dificuldade em encontrar uma posição única e definitiva não é um sinal de fraqueza da minha argumentação, mas sim um reflexo fiel da estrutura profunda do

problema. A linguística, como atividade intelectual e científica, enfrenta dilemas genuínos, e não meras confusões conceituais que poderiam ser dissolvidas com uma definição mais precisa.

O dilema estrutural que atravessa toda a disciplina pode ser formulado da seguinte maneira. A linguística precisa, para ser reconhecida como uma ciência, de um certo grau de rigor formal. Ela precisa de modelos precisos, de regras explícitas, de hipóteses testáveis e de métodos que possam ser replicados e avaliados por outros pesquisadores. É essa busca por rigor que a aproxima, em alguns aspectos, das ciências naturais. No entanto, ao mesmo tempo, a linguística precisa de uma imensa sensibilidade contextual para dar conta da natureza social, histórica e cultural da linguagem. Ignorar o contexto em que a língua é usada, as relações de poder que a atravessam, as identidades que ela constrói e a história que a molda seria produzir uma ciência empobrecida. Essa exigência de rigor formal de um lado e de sensibilidade contextual do outro puxa a disciplina em direções opostas. E o que vemos na história da linguística são diferentes escolas teóricas tentando resolver essa tensão, cada uma à sua maneira, privilegiando um dos polos. Até hoje, nenhuma delas conseguiu resolver o dilema de forma completamente satisfatória.

Diante dessa constatação, minha proposta é que abandonemos a busca quixotesca por uma solução definitiva e única. Em vez disso, sugiro que adotemos o que estou chamando de pluralismo metodológico. A ideia central, inspirada em parte na filosofia anarquista da ciência de Paul Feyerabend (1975) e em parte na prática concreta de linguistas como William Labov (1972), é a de que não existe um único método correto para fazer linguística. Existem, isso sim, múltiplos métodos, cada um deles legítimo dentro de seu domínio de aplicação, e cada um com vantagens específicas para determinados tipos de problema.

Se o meu objetivo é investigar as propriedades mais abstratas da sintaxe, a estrutura profunda e universal que supostamente estaria na base de todas as línguas humanas, então os métodos da tradição gerativa, baseados em julgamentos de gramaticalidade de falantes idealizados, podem ser os mais apropriados. Se, por outro lado, meu interesse é entender como a língua varia de acordo com a classe social, a idade ou o gênero dos falantes, ou como ela muda ao longo do tempo, então os métodos quantitativos da sociolinguística, baseados na análise de grandes *corpora* de fala espontânea, serão muito mais adequa-

dos. Se, ainda, minha questão é sobre os processos cognitivos em tempo real que acontecem quando compreendemos ou produzimos uma sentença, então os experimentos controlados da psicolinguística, com suas medições precisas de tempo de reação e movimentos oculares, tornam-se indispensáveis. Esses diferentes métodos não são redutíveis uns aos outros. Uma regra sintática não pode ser "traduzida" sem perdas em um padrão de variação sociolinguística, nem um processo cognitivo pode ser integralmente capturado por um experimento de laboratório. Mas eles também não são completamente incomensuráveis. Eles se sobrepõem parcialmente, conversam entre si, e uma boa teoria linguística deve ser capaz de integrar as descobertas provenientes de todas essas fontes, construindo um quadro mais rico e complexo do fenômeno.

O pluralismo metodológico que defendo, portanto, evita dois extremos igualmente perigosos. De um lado, ele rejeita o fundacionismo, a ideia de que existe um único método fundamental que deveria ser o padrão-ouro para toda e qualquer investigação linguística. De outro, ele também rejeita o relativismo ingênuo, a ideia de que todos os métodos são igualmente bons para todos os propósitos, ou que a escolha entre eles é puramente arbitrária. A escolha do método mais adequado depende intrinsecamente dos propósitos práticos da investigação e do contexto específico em que ela se insere. E essa escolha não é arbitrária porque pode ser avaliada de forma pragmática. Podemos e devemos examinar quais métodos produzem as predições mais precisas, quais geram descobertas que podem ser replicadas independentemente, quais dão origem a teorias mais férteis e heuristicamente poderosas. O linguista reflexivo, nessa perspectiva, é aquele que não apenas aplica um método de forma automática, mas que é capaz de explicitar qual método está usando, por que o escolheu e quais são suas limitações.

Desse pluralismo metodológico decorrem algumas implicações práticas importantes para a formação e a prática do linguista. Uma delas é a necessidade de transdisciplinaridade, dialogando produtiva e criticamente com a psicologia cognitiva, para entender os mecanismos de processamento da linguagem, com a antropologia para compreender a função social e cultural da fala, com a sociologia para analisar as relações de poder que se manifestam na língua, com a filosofia para articular e examinar seus próprios pressupostos metateóricos, e com a ciência da computação para desenvolver modelos formais cada vez mais

sofisticados e para lidar com a explosão de dados digitais. Nenhuma disciplina, sozinha, tem o quadro completo.

E para construir esse quadro, os linguistas têm de se tornar mais explícitos sobre seus próprios compromissos metateóricos. Em vez de fingir uma neutralidade que não existe, é mais produtivo e honesto reconhecer que toda descrição linguística, por mais neutra que pareça, carrega consigo pressupostos ontológicos sobre o que a língua é, epistemológicos sobre como podemos conhecê-la e éticos sobre como devemos tratar os falantes. Trazer esses pressupostos para a superfície e submetê-los ao debate é uma forma de fortalecer a disciplina, e não de enfraquecê-la.

Em suma, o pluralismo metodológico não é uma fórmula mágica que vai resolver todos os dilemas da linguística de uma vez por todas. O que ele oferece é uma forma produtiva de conviver com essas tensões, reconhecendo que elas não são defeitos a serem eliminados, mas sim características constitutivas de uma disciplina que se debruça sobre um objeto tão fascinante e multifacetado quanto a linguagem humana. Uma linguística verdadeiramente madura, sugiro, é aquela que abraça essas tensões em vez de tentar suprimi-las com reducionismos apressados ou falsas promessas de unificação.

6. Considerações Finais

Ao longo deste capítulo, empreendemos uma jornada por três dos problemas filosóficos mais centrais para a linguística. Começamos indagando sobre a natureza da língua. O que é, afinal, essa entidade que estudamos? Vimos que as respostas variam desde o platonismo, que a vê como um objeto abstrato, até o behaviorismo, que a reduz a comportamento, passando pelo psicologismo, que a localiza na mente individual, e pelo institucionalismo, que a entende como uma convenção social. Nenhuma dessas respostas se mostrou plenamente satisfatória, e a saída que propusemos foi a de um realismo linguístico modesto, que reconhece a realidade das múltiplas dimensões (cognitiva, social, formal) que constituem o fenômeno linguístico. Em seguida, investigamos a possibilidade e os métodos do conhecimento linguístico, i.e., como podemos conhecer as línguas e justificar nossas teorias sobre elas. Vimos que diferentes tradições metodológicas privilegiam diferentes tipos de evidência, cada uma com suas

virtudes e seus vícios, e que a escolha entre teorias empiricamente equivalentes nos força a recorrer a virtudes teóricas e a exercer um julgamento informado. Diante da complexidade, defendemos um pluralismo metodológico que integre, em vez de excluir, as múltiplas fontes de evidência. Por fim, nos voltamos para as responsabilidades éticas dos linguistas, e nos deparamos com dilemas profundos envolvendo justiça linguística, documentação de línguas ameaçadas e o papel social do pesquisador. A saída, argumentei, está em uma ética da colaboração, do respeito à autonomia e da reflexividade.

Essa conclusão tem implicações diretas para a maneira como concebemos a relação da linguística com outras áreas do conhecimento. A linguística não pode, e não deve, pretender ser autossuficiente. Para avançar em sua compreensão da linguagem, ela precisa dialogar profunda e continuamente com as ciências cognitivas, para desvendar os mecanismos de processamento e aquisição da linguagem; com as ciências sociais, para entender as dimensões históricas, políticas e culturais da vida linguística; e com a filosofia, para articular, examinar e, quando necessário, revisar os compromissos ontológicos, epistemológicos e éticos que estruturam sua própria prática.

A filosofia da linguística, portanto, é uma parte constitutiva e essencial da prática linguística reflexiva. Todo linguista, em seu trabalho cotidiano, ao escolher um método, ao formular uma hipótese, ao interpretar um dado, ao escrever um artigo, está, ainda que implicitamente, fazendo escolhas filosóficas. A diferença entre um bom linguista e um excelente linguista pode estar justamente na capacidade de tornar essas escolhas explícitas, de examiná-las criticamente e de justificá-las de forma coerente. Ser consciente dos compromissos filosóficos que estruturam a pesquisa é muito mais produtivo do que ser ignorante deles.

A linguística, vista sob essa perspectiva, revela-se um dos campos mais filosoficamente ricos e desafiadores de todo o espectro das ciências humanas. Suas tensões internas, que para alguns poderiam parecer fraquezas, são, na verdade, sua maior força. São elas que nos forçam a uma reflexão contínua e inescapável sobre o que significa, afinal, estudar a linguagem, esse fenômeno que é, ao mesmo tempo, um sistema cognitivo individual, uma instituição social coletiva, um objeto formal de investigação científica e a própria matéria-prima da nossa humanidade.

Este capítulo, portanto, deu continuidade à jornada mais ampla que iniciamos com a antropologia. Lá, aprendemos que a cultura é o meio através do qual os humanos criam e partilham significados. Aqui, na linguística, aprendemos que a linguagem é o veículo fundamental através do qual esse significado é transmitido, preservado e transformado. A linguagem é o que torna a cultura algo que pode ser compartilhado entre as pessoas, ensinado às novas gerações e herdado do passado. É ela que dá forma ao pensamento e que nos permite construir, coletivamente, os mundos simbólicos em que habitamos.

No entanto, a linguagem falada, por mais poderosa que seja, tem uma limitação crucial, pois é efêmera. As palavras se dissipam no ar assim que são pronunciadas. Para que o significado cultural criado pela antropologia e transmitido pela linguagem possa durar, para que possa atravessar o tempo e chegar às gerações futuras, ele precisa se tornar registro. Precisa se fixar em um suporte material que o preserve. Antes de chegarmos a esse momento do registro, porém, somos convidados a fazer uma escala em uma forma ainda mais fundamental de abstração e linguagem, a matemática. Se a linguagem é o veículo do significado cultural, a matemática pode ser vista como a linguagem da própria razão, aquela que parece descrever verdades eternas e universais, que existiriam independentemente de qualquer cultura ou época específicas. É exatamente esse o território que exploraremos no próximo capítulo, dedicado à Filosofia da Matemática, onde a tensão entre o abstrato e o concreto, entre o universal e o particular, entre o descoberto e o inventado, atingirá o seu ponto mais agudo e fascinante.

4

Filosofia da Matemática

1. Por que uma Filosofia da Matemática?

A matemática é estranha. Penso nisso com frequência. Dois mais dois é igual a quatro. Sempre foi, sempre será. O Teorema de Pitágoras era verdadeiro antes de existirem humanos para pensá-lo e continuará verdadeiro muito depois que nossa espécie desaparecer. A infinitude dos números primos é uma verdade que parece não depender de nada físico, de nenhum experimento, de nenhuma observação do mundo. A matemática tem essa aura de certeza absoluta, de necessidade, de eternidade. Como isso é possível?

Essa estranheza é o ponto de partida da filosofia da matemática; ela não é apenas uma perplexidade inicial que a filosofia dissipa, mas um fenômeno filosófico genuíno que resiste a explicações simples. A necessidade aparente das verdades matemáticas, sua independência da experiência e sua aplicabilidade ao mundo físico constituem o que Hilary Putnam (1979) chamou de "o problema da verdade matemática", um conjunto de questões que qualquer teoria filosófica da matemática deve enfrentar (Putnam, 1979, p. 60-78).

Essa estranheza fica mais clara quando comparamos a matemática com outras ciências. A física descobre leis sobre a matéria e a energia. A biologia descobre padrões da vida. Essas descobertas dependem crucialmente de observação, experimento, evidência empírica coletada no mundo. Mas o matemático não observa nada. Ele prova teoremas sentado em uma cadeira, pensando, raciocinando a partir de definições e axiomas. Não faz experimentos em laboratório, não coleta dados. Apenas pensa. E, no entanto, descobre verdades que parecem mais certas, mais robustas, mais eternas do que qualquer verdade empírica. A física newtoniana foi substituída pela de Einstein, mas o teorema de Pitágoras continua lá, inabalado.

Isso levanta questões filosóficas profundas, daquelas que nos fazem coçar a cabeça. Sobre o que é a matemática, afinal? Os números existem? Se sim, onde eles estão? Eles habitam algum reino platônico de ideias abstratas? Se não, como pode a matemática ser sobre alguma coisa? Como podemos conhecer verdades matemáticas com tamanha certeza? Por que a matemática, algo aparentemente tão abstrato, se aplica tão perfeitamente ao mundo físico, como Eugene Wigner (1960) chamou de "a eficácia irrazoável da matemática nas ciências naturais"? E há também questões éticas. Os matemáticos têm responsabilidades sobre como suas descobertas abstratas são usadas no mundo real? A criptografia que protege nossos dados, os algoritmos que governam a inteligência artificial, tudo isso depende de matemática. Se a matemática causa danos, de quem é a culpa?

Essas três ordens de questões, metafísica (o que são os objetos matemáticos), epistemológica (como os conhecemos) e ética (que responsabilidades decorrem do conhecimento matemático), não são independentes. A resposta que se dá à questão metafísica sobre o que são os números determina, em grande medida, o que contará como conhecimento legítimo em matemática; e a concepção do conhecimento matemático, por sua vez, delimita o escopo das responsabilidades dos matemáticos. A articulação dessas três dimensões é o fio condutor deste capítulo.

Este capítulo ocupa uma posição estratégica na jornada que estamos percorrendo. Viemos da antropologia, que nos mostrou que a cultura é o meio pelo qual os humanos criam significado. Passamos pela linguística, que nos mostrou que a linguagem é o veículo pelo qual esse significado é transmitido. Agora, a matemática nos mostra algo ainda mais fundamental, a saber, a estrutura abstrata que subjaz tanto à linguagem quanto à cultura, e que torna possível a própria ideia de informação, de computação, de criptografia. A matemática é, de certa forma, a linguagem da própria razão. E, como veremos em capítulos futuros, sobre ciência da informação, ciência da computação e criptografia, ela é também a linguagem que torna possível o mundo digital. Compreender sua filosofia é, portanto, compreender as fundações de tudo o que virá.

Não vou fingir que tenho respostas definitivas. Os problemas filosóficos sobre a matemática são antigos, discutidos desde Platão (2001), e ainda não foram resolvidos de forma consensual. Mas posso tentar articular os problemas

com alguma precisão, mapear as principais posições, avaliar seus méritos e fraquezas. E, ao final, propor o que me parece a posição mais defensável, embora reconhecendo suas limitações. É um exercício de honestidade intelectual.

O dilema central da filosofia da matemática é uma tensão entre o platonismo e o nominalismo. O platonismo explica muito bem a objetividade, a necessidade e a eternidade da matemática, mas postula a existência de um reino de entidades abstratas que são metafisicamente misteriosas e apartadas do espaço e do tempo. Como podemos conhecer algo que não está no espaço nem no tempo? O nominalismo, por sua vez, resolve esse mistério metafísico, negando a existência dessas entidades, mas não consegue explicar de forma satisfatória por que a matemática funciona tão bem e por que ela parece tão objetiva e não arbitrária. As posições intermediárias que exploraremos são tentativas de navegar entre esses dois polos, cada uma pagando um preço filosófico diferente.

Vale notar, desde o início, que esse debate não é meramente acadêmico; ele tem consequências práticas para o modo como os matemáticos concebem seu trabalho, para o modo como a matemática é ensinada e para o modo como se avalia a responsabilidade ética de quem desenvolve matemática com aplicações poderosas. Uma filosofia da matemática não é um ornamento cultural; é uma condição de inteligibilidade da prática matemática.

2. Metafísica da Matemática

A questão metafísica fundamental é a mais simples de formular, mas a mais difícil de responder. Os números existem? Se sim, onde e como? Se não, sobre o que, afinal, estamos falando quando fazemos afirmações matemáticas?

O platonismo é a posição mais antiga e, desconfio, a mais intuitiva para a maioria dos matemáticos praticantes. Platão (2001) argumentava que, além do mundo físico e mutável dos objetos que percebemos com os sentidos, existe um mundo de formas ou ideias eternas, perfeitas e imutáveis, somente apreensível pela mente. Os círculos que desenhamos na areia são imperfeitos e apagam-se, mas a forma do círculo é perfeitamente circular e eterna. Da mesma forma, o número três em si não são os conjuntos de três objetos que vemos por aí, mas a própria forma do três, da qual esses conjuntos participam. A matemática seria, então, o estudo desse reino abstrato.

O platonismo em matemática não é monolítico; há versões que diferem quanto ao tipo de acesso que temos ao reino abstrato, quanto à sua relação com o mundo físico e quanto ao escopo dos objetos abstratos que postulam. Frege (1983) e Gödel (1989) defenderam versões distintas, como veremos a seguir, e versões mais recentes, como o platonismo plenitudinário de Mark Balaguer (1998), tentam responder às objeções clássicas de formas inovadoras (Balaguer, 1998, p. 35-64).

Na filosofia contemporânea, o platonismo foi defendido com rigor por Gottlob Frege (1983) e Kurt Gödel (1989); e, no caso das leis da natureza, eu mesmo defendo uma forma de platonismo heterodoxo. Frege (1983) argumentou que os números não são propriedades de objetos físicos (como "três laranjas") nem meras construções mentais. Eles são, na verdade, objetos lógicos, autônomos e abstratos. O número três é um objeto abstrato que pode ser "exemplificado" por trios de coisas no mundo. Já Gödel (1989), um dos maiores lógicos do século XX, foi ainda mais longe. Ele acreditava que temos uma intuição matemática, uma espécie de percepção intelectual que nos coloca em contato direto com esse mundo de objetos abstratos, da mesma forma que a percepção sensorial nos coloca em contato com o mundo físico.

A posição de Frege (1983) é especialmente importante porque emerge de sua tentativa de fundamentar a aritmética na lógica pura, o projeto logicista. Para Frege (1983), que os números sejam objetos lógicos não é uma hipótese metafísica extravagante, mas uma consequência da análise do conteúdo de nossas afirmações aritméticas. A crise gerada pelo paradoxo de Russell, que mostrou que o sistema lógico de Frege (1983) era inconsistente, não refuta a intuição subjacente de que os números são objetivos e independentes da mente, mas obriga a uma reformulação do modo como os fundamentamos (Frege, 1884/1980, p. 67-99).

O platonismo tem inúmeras vantagens, como a de explicar por que a matemática é objetiva, necessária e conhecível *a priori*. "Dois mais dois é quatro" não é verdadeiro porque alguém decidiu que sim, mas porque é um fato sobre objetos abstratos independentes de nós. E tais proposições não poderiam ser diferentes porque são sobre estruturas abstratas que existem necessariamente da forma que são. Como a matemática é sobre tal reino abstrato, não sobre o mundo físico, não precisamos observar o mundo para conhecê-la. Além disso, o

platonismo dá conta da nossa intuição de que quando um matemático prova um novo teorema, ele não está criando algo novo, mas descobrindo uma verdade que já estava lá, esperando para ser encontrada.

Para avaliar bem o platonismo, não podemos apenas nos enfocar nas vantagens, mas temos de atentar também para as suas desvantagens. Uma delas é que não está claro onde se encontra esse reino de objetos abstratos. Certamente não no espaço físico, porque são abstratos. Não no tempo, porque são eternos. Não em mentes individuais, porque são objetivos e independentes. Então onde? Platão (2001) falava de um "mundo das formas", mas isso não é uma explicação, é uma metáfora. Uma ontologia de objetos fora do espaço e do tempo é profundamente misteriosa. E, se não dissolvermos esse mistério, isso permanece como uma desvantagem teórica.

Paul Benacerraf (1973) traz ao debate o problema epistemológico do platonismo. Se os objetos matemáticos são abstratos, fora do espaço e do tempo, então eles não podem interagir causalmente conosco. Não podemos vê-los, tocá-los, detectá-los de forma alguma. Mas então como os conhecemos? O conhecimento empírico padrão parece requerer uma relação causal entre o conhecedor e o objeto conhecido. Vejo uma mesa porque a mesa causa certas impressões em minha retina. Ouço um som porque as ondas sonoras causam vibrações em meu tímpano. Mas números não causam nada. Logo, se o platonismo estiver correto, parece que não podemos conhecer verdades matemáticas. Mas obviamente conhecemos. Então há uma dificuldade aqui.

O argumento de Benacerraf (1973) tem a forma de um dilema. Ou aceitamos uma semântica uniforme para enunciados matemáticos e não matemáticos, em que ambos são verdadeiros em virtude de relações com objetos, e então enfrentamos o problema epistemológico do platonismo; ou adotamos uma semântica diferente para enunciados matemáticos que dissolva o problema epistemológico, mas então sacrificamos a uniformidade semântica e a objetividade da matemática. Nenhuma posição filosófica escapou completamente a esse dilema (Benacerraf, 1973, p. 661-679).

Para solucionar esse desafio, Gödel (1989) propôs que temos uma faculdade especial da intuição matemática, análoga à percepção sensorial, mas direcionada a objetos abstratos. Mas isso é obscuro. Que tipo de processo cognitivo é esse? Como funciona? Por que deveríamos confiar nele? Não temos nenhuma

resposta. Outros apelaram ainda à necessidade conceitual ou à inferência para a melhor explicação. Mas nenhuma resposta é totalmente convincente. O problema epistemológico para o platonismo permanece.

Ele é também extravagante ontologicamente. O platonismo postula um número infinito de objetos abstratos, a saber, todos os números naturais, reais, complexos, conjuntos, funções, estruturas matemáticas concebíveis. Isso viola o princípio de parcimônia ontológica. Não deveríamos multiplicar entidades além da necessidade. E talvez a matemática possa ser explicada sem todo esse aparato metafísico.

E foi isso que os oponentes dos platonistas, os nominalistas, fizeram. Negaram a existência de objetos abstratos. Os nominalistas argumentam que números não existem como entidades independentes; existem apenas os objetos físicos concretos no espaço e no tempo. A matemática não é sobre um reino especial de objetos abstratos, mas sobre símbolos, convenções linguísticas, ou padrões em objetos concretos. Quando dizemos que "dois mais dois é quatro", não estamos descrevendo uma relação entre entidades abstratas, mas simplesmente seguindo regras de uma linguagem que criamos.

O nominalismo contemporâneo em filosofia da matemática articula-se com o nominalismo clássico de Guilherme de Ockham, para quem apenas existem indivíduos particulares e os universais são nomes que aplicamos a conjuntos de particulares. Transposto para a matemática, isso significa que "três" não nomeia um objeto abstrato, mas funciona como um predicado ou um operador aplicado a conjuntos de objetos concretos. O desafio central para o nominalista é explicar como enunciados matemáticos podem ser verdadeiros.

Hartry Field (1980) desenvolveu uma versão sofisticada do nominalismo mostrando como reconstruir a física sem usar a matemática. A física usual quantifica sobre números, funções, espaços vetoriais. Field (1980) mostrou como reformulá-la usando apenas quantificação sobre regiões espaço-temporais concretas. Isso prova, segundo ele, que a matemática é dispensável para a ciência; logo, não precisamos acreditar em objetos matemáticos. Eles são ficções úteis, não entidades reais.

O programa de Field (1980) é tecnicamente ambicioso, mas enfrenta objeções sérias além da complexidade. Uma delas, levantada por Shapiro (1997), é que a reformulação nominalista da física pressupõe ela mesma uma ontologia

de regiões espaço-temporais que pode ser tão problemática quanto a ontologia de objetos matemáticos. Outra objeção é que a ficcionalidade da matemática não explica por que as ficções matemáticas são indispensáveis para a prática científica; se são ficções, por que funcionam tão bem? (Field, 1980, p. 1-50; Shapiro, 1983, p. 521-544).

O nominalismo tem virtudes importantes, como a parcimônia ontológica, por não postular entidades misteriosas fora do espaço-tempo e por se comprometer apenas com o mundo físico investigável pelas ciências naturais. E tem também algumas vantagens, como evitar o problema epistemológico de Benacerraf (1973); afinal, se números não existem, não precisamos explicar como os conhecemos; precisamos apenas explicar como usamos a linguagem matemática, o que é uma questão bem mais tratável.

Mas o nominalismo também tem custos pesados. Um deles é o fato de que ele não explica a objetividade e nem a aplicabilidade da matemática. Se os números são ficções ou convenções, então "dois mais dois é quatro" parece análogo a "Sherlock Holmes mora em Baker Street". Mas isso soa errado. Verdades matemáticas parecem necessárias de uma forma que verdades ficcionais não são. Os nominalistas podem dizer que são necessárias dadas as convenções que adotamos, mas isso ainda parece não capturar a objetividade forte que a matemática exibe. Ademais, se a matemática é apenas sobre convenções ou símbolos, por que ela funciona tão bem para descrever o mundo físico? Wigner (1960), conforme mencionamos, chamou isso de "a eficácia irrazoável da matemática". Físicos descobrem que equações matemáticas desenvolvidas por razões puramente internas à matemática preveem fenômenos físicos com precisão impressionante. Se a matemática não é sobre a realidade, isso é uma coincidência milagrosa. Os platonistas têm um caminho de explicação; eles nos dizem que a matemática descreve as estruturas abstratas que o mundo físico exemplifica ou instancia. Os nominalistas não têm nenhuma explicação comparável.

Além disso, o programa de Field (1980) de eliminar a matemática da física é tecnicamente complexo e pouco usado por cientistas praticantes. Se precisamos da matemática para fazer ciência *eficientemente*, então, independentemente do status ontológico, a matemática é indispensável *pragmaticamente*. Isso

não diz que o nominalismo está errado, mas enfraquece significativamente a motivação para sustentá-lo.

Como caminhos alternativos, há algumas posições intermediárias, que tentam capturar virtudes de ambos os lados, evitando seus vícios. O formalismo, associado a David Hilbert (1926/1967), por exemplo, diz que a matemática não é sobre objetos, mas sobre a manipulação de símbolos seguindo regras. A matemática é um jogo formal. Os números não existem como objetos, mas os símbolos numerais existem, e as regras determinam como manipulá-los. As verdades matemáticas são verdades sobre quais expressões são deriváveis de axiomas, dadas certas regras de inferência.

O formalismo de Hilbert (1926/1967) foi motivado pela crise dos fundamentos gerada pelas antinomias da teoria ingênua dos conjuntos, em particular o paradoxo de Russell (1902) e o paradoxo de Burali-Forti (1897). O programa de Hilbert (1926/1967) visava mostrar que a matemática poderia ser reduzida a sistemas formais cuja consistência seria provada por métodos finitários, i.e., por métodos que não pressupõem o infinito atual. Era uma tentativa de dar à matemática uma fundação segura, imune às antinomias, sem precisar de objetos platônicos. O teorema da incompletude de Gödel (1931) destruiu essa esperança, ao demonstrar que nenhum sistema formal consistente e suficientemente poderoso pode provar sua própria consistência (Hilbert, 1926/1967, p. 367-392; Gödel, 1931/1967, p. 596-616).

O formalismo evita o mistério ontológico do platonismo, sem negar completamente a matemática, como o nominalismo às vezes parece fazer. A matemática não é sobre nada além de símbolos, mas os símbolos são concretos, isto é, podem ser escritos no papel ou armazenados em computadores. A verdade matemática é uma questão de derivabilidade formal. O problema desse caminho é justamente que as regras matemáticas seriam arbitrárias; e isso é contrário ao que de fato parece ser o caso, a saber, que as verdades matemáticas são descobertas, e não inventadas. Além disso, se a matemática é apenas um jogo formal, com regras arbitrárias, por que as aplicações ao mundo físico funcionam? Não pode ser mera coincidência que um jogo formal seja tão extraordinariamente útil para a física.

Além disso, o formalismo enfrenta a questão de como escolher os axiomas. Se a matemática é apenas um jogo formal, então qualquer conjunto de regras é

tão bom quanto qualquer outro. Mas os matemáticos não pensam assim. Eles consideram que certos axiomas são mais "verdadeiros" ou mais "naturais" que outros. O axioma da escolha é controverso precisamente porque não é intuitivamente óbvio, mas tem consequências poderosas. Isso sugere que há mais na matemática do que meras regras formais.

O intuicionismo, desenvolvido por L.E.J. Brouwer (1913/1983), oferece uma alternativa diferente. A matemática não é sobre objetos abstratos platônicos nem sobre símbolos formais, mas sobre construções mentais. Objetos matemáticos são criados pela atividade mental dos matemáticos. A existência matemática é inseparável da possibilidade de construção por um sujeito ideal. A verdade matemática não é a correspondência com uma realidade abstrata, mas a construtibilidade, i.e., uma proposição é verdadeira se e somente se temos uma construção mental que a estabelece.

O intuicionismo de Brouwer (1913/1983) tem raízes na tradição kantiana. Assim como Kant (2001) argumentou que a matemática é possível porque espaço e tempo são formas *a priori* da intuição do sujeito, Brouwer (1913/1983) argumenta que a matemática é uma atividade mental que não depende de objetos externos, nem platônicos nem físicos. A diferença crucial é que Brouwer (1913/1983) recusa a lógica clássica como ferramenta da matemática; a lei do terceiro excluído (toda proposição é verdadeira ou falsa) não vale intuicionisticamente, porque uma proposição só é verdadeira se construída e só é falsa se refutada, e há proposições para as quais não temos nem um nem outro (Brouwer, 1913/1983, p. 90-96).

O intuicionismo tem um apelo considerável, porque, nele, conhecemos os objetos matemáticos diretamente, já que os construímos, solucionando assim o problema epistêmico, sem a postulação ontológica de um reino para as *abstracta*, dado que os objetos matemáticos são construções mentais. Mas tal teoria também tem consequências radicais para a lógica, dado que nem toda proposição vai ser verdadeira ou falsa, pelo fato de uma proposição só ser verdadeira se construirmos uma prova e só ser falsa se construirmos uma refutação; se não fizemos nem um nem outro, ela não é nem verdadeira nem falsa. Isso força a revisão da lógica clássica e de grande parte da matemática tradicional. Muitas provas por contradição não são válidas intuicionisticamente, e alguns teoremas clássicos deixam de ser teoremas. Para os matemáticos, isso é disruptivo de-

mais. Eles preferem manter a matemática clássica. E filosoficamente, o intuicionismo enfrenta um dilema sério na sua ideia de matemática como construção mental, a saber, se as construções mentais são privadas a cada indivíduo, perdemos a objetividade da matemática, e se são objetivas, elas pareceriam ser abstratas novamente.

O estruturalismo, desenvolvido por Benacerraf (1965, 1973) e Shapiro (1997), oferece outra via. A matemática não é sobre objetos, mas sobre estruturas. Não importa o que os números são intrinsecamente; importa apenas como se relacionam estruturalmente. O número três é uma posição na progressão dos números naturais, definida por suas relações com outros números. A estrutura é o que importa, não a natureza intrínseca dos ocupantes das posições.

O estruturalismo captura bem porque múltiplas construções conjuntistas dos números naturais, a construção de von Neumann, a construção de Zermelo. são igualmente legítimas, pois elas realizam a mesma estrutura. O que importa para a aritmética não é qual delas representa os números de verdade, mas que ambas instanciam a estrutura dos naturais. Shapiro (1997) distingue o estruturalismo *ante rem*, segundo o qual as estruturas existem independentemente de qualquer instância, e o *in re*, segundo o qual estruturas apenas existem enquanto instanciadas em sistemas. O primeiro é uma forma de platonismo de estruturas; o segundo enfrenta dificuldades para acomodar estruturas matemáticas sem instâncias físicas (Shapiro, 1997, p. 71-107).]

O estruturalismo captura bem a prática matemática moderna. Os matemáticos raramente se preocupam com o que os números "são" em si mesmos, atentando apenas para suas propriedades estruturais. A teoria dos grupos estuda estruturas grupais abstratamente, não grupos específicos de objetos. A topologia estuda propriedades estruturais de espaços, não espaços particulares. O problema então é que não está claro o que são as estruturas. Se são objetos abstratos, voltamos ao platonismo; se são construções ou convenções, voltamos aos problemas do nominalismo ou do formalismo. E, ainda, parece que se números são meras posições em estruturas, sem natureza intrínseca, então Júlio César poderia ser o número três. Se dissermos que não, como obviamente parece ser o caso, o estruturalismo teria dificuldades em explicar a razão.

Face às dificuldades de cada uma dessas posições, proponho o que chamo de naturalismo matemático modesto, inspirado em filósofos como Quine e Pene-

lope Maddy (1997). Esta posição tenta preservar intuições centrais de cada lado sem cair em seus problemas mais graves. O naturalismo matemático modesto sustenta que a matemática é contínua com a ciência empírica, e não radicalmente separada. Verdades matemáticas estão na mesma rede de crenças que as verdades físicas, embora estejam mais centrais na rede e sejam menos revisáveis. O compromisso ontológico com objetos matemáticos seria justificado pragmaticamente pela indispensabilidade da matemática para a ciência. Se precisamos da matemática para fazer física, então estamos justificados em acreditar na existência de objetos matemáticos, mesmo que sua natureza precise permanecer misteriosa.

Maddy (1997) desenvolve uma versão refinada do naturalismo que vai além de Quine (1960) em um ponto crucial; em vez de justificar a crença em objetos matemáticos pela indispensabilidade para a física, ela sustenta que a matemática deve ser avaliada segundo seus próprios padrões metodológicos internos, sem redução a outras ciências. Os matemáticos têm autoridade para decidir quais métodos e axiomas são aceitáveis dentro da matemática, e o filósofo deve respeitar essa autoridade em vez de impor critérios externos. Isso é o que Maddy (1997) chama de "naturalismo matemático" em sentido estrito (Maddy, 1997, p. 143-202).

Nesse sentido, a questão sobre "onde" os objetos matemáticos existem é mal formulada. Objetos abstratos não existem no espaço e no tempo, mas isso não significa que não existam. Significa apenas que seu modo de existência é diferente do modo de existência de objetos físicos. Ele é conhecido *a priori* no sentido de que não depende de experimentos específicos, mas é *a posteriori* no sentido de que depende da prática matemática, que é uma atividade humana situada historicamente. Não há uma faculdade especial da intuição matemática, apenas o raciocínio humano comum aplicado a domínios abstratos.

Assim, a aplicabilidade da matemática não é um milagre, mas uma consequência de que a matemática e a física evoluíram juntas. A matemática foi desenvolvida, em grande parte, para resolver problemas práticos e teóricos colocados pela ciência. Portanto, não é surpreendente que ela se aplique bem. Da mesma forma, quando um economista cria um modelo econômico para certo tipo de situação real, não é impressionante que ele se adeque bem à realidade, ainda que o modelo seja composto apenas por objetos abstratos.

Esta posição evita o mistério do platonismo robusto e do intuicionismo, sem cair no ceticismo do nominalismo. Pois, ainda que a ontologia matemática seja problemática, esse é um problema com o qual podemos conviver, dada a utilidade da matemática. Isso está diretamente alinhado com a prática real dos matemáticos, que raramente se preocupam com questões ontológicas enquanto fazem matemática.

Como essa posição se conecta com capítulos anteriores? Na antropologia, vimos que a cultura é real, mas seu modo de existência é complexo. Na linguística, vimos que línguas são simultaneamente cognitivas, sociais e abstratas. Na matemática, vemos o mesmo padrão. Os objetos matemáticos são reais, mas seu modo de existência é peculiar. O pluralismo ontológico que emerge aqui ecoa o pluralismo dos capítulos anteriores e prepara o terreno para a ciência da informação, onde a informação também terá natureza dual, física e semântica, abstrata e concreta.

Esta posição evita o mistério do platonismo robusto e do intuicionismo, sem cair no ceticismo do nominalismo. Ainda que a ontologia matemática seja problemática, trata-se de um problema com o qual podemos conviver, dada a utilidade da matemática. Essa é, precisamente, a lição do pragmatismo filosófico de William James (1890) e John Dewey (1938/1986), segundo a qual, quando a resolução de uma questão teórica não altera a prática, a questão pode ser suspensa sem que isso comprometa a racionalidade da prática. Aplicado à filosofia da matemática, isso significa que podemos fazer matemática rigorosa sem decidir primeiro se os números existem, mas não podemos ignorar indefinidamente a questão das responsabilidades éticas decorrentes da prática matemática.

3. Epistemologia da Matemática

A metafísica pergunta o que são objetos matemáticos. A epistemologia pergunta como os conhecemos. Essas questões estão conectadas, mas são distintas. Mesmo que resolvêssemos a metafísica definitivamente, a epistemologia permaneceria desafiadora. A distinção tradicional entre conhecimento *a priori* e *a posteriori* é central aqui. O conhecimento *a posteriori* depende da experiência sensorial. Sabemos que neve é branca porque vemos a neve branca. O conhecimento *a priori*, por sua vez, não depende da experiência. Sabemos que solte-

ros são não-casados sem observar nenhum solteiro. A matemática é tradicionalmente o paradigma do conhecimento *a priori*. Sabemos que dois mais dois é quatro sem contar objetos. Apenas pensamos e vemos que deve ser assim. A distinção *a priori*/ *a posteriori* é, ela própria, filosoficamente disputada. Kripke (1980) mostrou que ela não coincide com a distinção necessário/contingente, pois há verdades necessárias conhecidas *a posteriori* (por exemplo, que a água é H_2O) e, possivelmente, verdades contingentes conhecidas *a priori*. Isso complica o quadro tradicional em que a matemática é necessária e *a priori* por contraste com a ciência empírica, que é contingente e *a posteriori*. A epistemologia da matemática precisa navegar por essas distinções sem presumir que elas são mais simples do que são (Kripke, 1980, p. 34-60).

Kant (2001) articulou isso classicamente. Juízos matemáticos são sintéticos *a priori*. Sintéticos porque não são meras verdades por definição; eles expandem nosso conhecimento. E *a priori* porque não dependem da experiência. Kant (2001) argumentou que a matemática é possível porque o espaço e o tempo são formas *a priori* da intuição. A geometria euclidiana descreve a estrutura do espaço como inevitavelmente o experienciamos. A aritmética descreve a estrutura de nossa intuição temporal.

Essa explicação foi abalada por desenvolvimentos posteriores. A descoberta de geometrias não-euclidianas mostrou que a geometria euclidiana não é a única possibilidade. A teoria da relatividade mostrou que o espaço físico não é euclidiano. Logo, a geometria não descreve a estrutura necessária da experiência possível; é uma teoria empírica sobre o mundo físico, testável e revisável. Parte da matemática, pelo menos, é *a posteriori*. Mas a aritmética parece diferente. Dois mais dois é quatro parece necessário de uma forma que o teorema de Pitágoras não é, agora que sabemos que o espaço pode ser não-euclidiano. Talvez a distinção não seja entre matemática e ciência empírica, mas dentro da própria matemática entre partes *a priori* e partes *a posteriori*.

A distinção entre geometria e aritmética quanto ao seu estatuto epistêmico é reveladora. Ela sugere que a matemática não é uma disciplina epistemicamente homogênea. Frege (1983) já havia argumentado que a aritmética é lógica, portanto analítica e *a priori*, enquanto a geometria seria sintética *a priori*. Após o colapso do logicismo fregeano com o paradoxo de Russell, e após a relativização da geometria pela física, a questão sobre quais partes da matemá-

tica são genuinamente *a priori* permanece em aberto, e a resposta pode variar conforme a interpretação metafísica adotada.

Os positivistas lógicos ofereceram uma explicação diferente. Carnap e Ayer (1936/1952) argumentaram que a matemática é analítica, i.e., verdadeira pelo significado dos termos e pelas regras lógicas. Dois mais dois é quatro porque é uma consequência lógica das definições de números e operações. Não precisamos de intuição especial ou faculdade *a priori*. Apenas precisamos entender os significados e seguir a lógica.

O argumento de Ayer (1936/1952) em favor da analiticidade da matemática baseia-se numa concepção deflacionária da necessidade. Dizer que uma proposição é necessariamente verdadeira é dizer que a negaríamos somente ao preço de revisar as definições dos termos envolvidos. Isso é uma versão linguística da necessidade, que a reduz a compromissos convencionais. A objeção de Quine (1951) à distinção analítico/sintético é precisamente que não há fundamento para distinguir revisões que alteram "significado" de revisões que alteram "fato", a distinção pressupõe exatamente o que precisaria ser demonstrado (Ayer, 1936/1952, p. 71-87; Quine, 1951, p. 20-43).

Essa visão tem elegância. Ela explica a necessidade da matemática e sua aprioridade, evitando o misticismo da intuição de abstratos. Mas enfrenta também um problema famoso, de se a matemática é analítica, ela não informa sobre a realidade. Verdades analíticas são vazias de conteúdo factual. Mas a matemática claramente informa sobre a realidade; ela é indispensável para a ciência. Logo, não pode ser puramente analítica.

Um caminho aqui seria o de Quine (1951), que atacou a própria distinção analítico-sintético. Ele argumentou que não há linha clara separando verdades por significado de verdades por fato. O conhecimento forma uma rede conectada. Evidência empírica afeta toda a rede holisticamente. A matemática está na rede, conectada à física e às outras ciências. Logo, é revisável empiricamente a princípio, embora esteja tão central na rede que seria a última a ser revisada. A matemática é contínua com a ciência empírica, não radicalmente diferente.

Se a matemática é revisável empiricamente, então ela pode estar errada. Descobertas empíricas poderiam nos forçar a revisar a lógica ou a aritmética. Isso parece absurdo, entretanto. Como poderia uma evidência empírica mostrar que dois mais dois não é quatro? Quine (1951) argumenta que, se o universo

fosse suficientemente estranho (talvez com um objeto surgindo sempre que juntamos quatro), poderíamos achar mais simples revisar a aritmética do que a física.

Na prática, nenhum matemático pensa assim. Matemáticos tratam provas como estabelecendo verdades com certeza, não como hipóteses testáveis. E há boas razões para isso. Provas matemáticas são dedutivamente válidas. Se os axiomas são verdadeiros e a lógica é sólida, a conclusão deve ser verdadeira. Não há espaço para erro empírico. Essa certeza distintiva da matemática precisa ser explicada. Talvez a distinção relevante não seja entre *a priori* e *a posteriori*, mas entre conhecimento por prova dedutiva e conhecimento por inferência indutiva. A matemática usa o primeiro. As ciências empíricas usam o segundo. Provas dedutivas garantem a verdade das conclusões, dadas as premissas. Inferências indutivas apenas tornam conclusões prováveis. Isso explicaria por que a matemática tem uma certeza que ciências empíricas não têm.

Mas isso apenas empurra a questão. De onde vêm os axiomas? Eles não são provados, são assumidos. Por que acreditamos neles? Se o platonismo é verdadeiro, os axiomas são verdades sobre abstratos que intuímos de alguma forma. Se o nominalismo é verdadeiro, axiomas são convenções que adotamos. Se o intuicionismo é verdadeiro, axiomas são construções mentais básicas. Cada metafísica sugere uma epistemologia diferente. Metafísica e epistemologia estão intimamente entrelaçadas.

Há também a questão de por que a lógica é confiável. Provas dependem de regras de inferência lógica. Mas como sabemos que a lógica é correta? Não podemos provar logicamente que a lógica é correta sem circularidade. Parece que temos que aceitar a lógica como básica, auto-evidente, não fundamentada em algo mais básico. Há um elemento não-racional no fundamento da racionalidade matemática. Isso é desconfortável.

Essa questão sobre a fundação da lógica é conhecida como o problema do regresso de Carroll (1895). Em *What the Tortoise Said to Achilles* (1895), Lewis Carroll (1895) mostrou que tentar justificar uma regra de inferência por meio de uma premissa adicional gera um regresso infinito. A conclusão filosófica é que pelo menos algumas regras de inferência devem ser aceitas como primitivas, não inferidas de nada mais básico. Isso não é uma fraqueza peculiar da matemática; é uma característica estrutural de qualquer sistema de conhecimento

que use inferência. Mas torna manifesto que a certeza matemática repousa, em seu fundamento, sobre compromissos que não são eles mesmos matematicamente justificados (Carroll, 1895, p. 278-280; Cid, 2016).

Matemáticos praticantes raramente se preocupam com essas questões. Eles fazem matemática confiando em intuições, provando teoremas, construindo teorias. A epistemologia filosófica parece irrelevante para a prática. Mas ocasionalmente surgem questões com dimensão epistemológica real. O axioma da escolha é controverso porque não é intuitivamente óbvio, mas tem consequências contraintuitivas. Grandes cálculos verificados por computador levantam a questão de se eles são provas genuínas, dado que nenhum humano pode verificá-las completamente.

A hipótese do contínuo não é provável nem refutável em ZFC. Isso levanta a questão de saber se ela é verdadeira ou falsa independentemente de provas. Platonistas diriam que sim, há um fato da matéria. Formalistas diriam que não, apenas há uma consequência dos axiomas. Essa não é questão puramente metafísica, pois afeta como matemáticos pensam sobre seu trabalho e que métodos usam.

O caso da hipótese do contínuo é particularmente instrutivo. Cohen (1963) provou que ela é independente dos axiomas de ZFC. Nem ela nem sua negação podem ser derivadas de ZFC se este for consistente. Isso gerou uma divisão entre matemáticos. Alguns, como Gödel (1989), acreditam que a hipótese tem um valor de verdade determinado que ultrapassa o que ZFC pode capturar, e que os axiomas mais fortes nos permitirão decidir a questão; outros concluem que a questão simplesmente não tem resposta determinada, o que é uma forma de pluralismo matemático. A escolha entre essas posições não é meramente técnica; ela reflete comprometimentos metafísicos sobre o que é a matemática (Cohen, 1963, p. 1143-1148).

Um dos problemas epistemológicos mais intrigantes sobre matemática é sua aplicabilidade. Wigner (1960) falou da “eficácia irrazoável da matemática nas ciências naturais”. Conforme dissemos, os físicos formulam teorias usando equações matemáticas desenvolvidas por razões puramente internas à matemática, e descobrem que essas equações descrevem o mundo físico com precisão impressionante. Por exemplo, a teoria dos grupos foi desenvolvida no século XIX como teoria matemática puramente abstrata. No século XX, tornou-se

fundamental para a física quântica. As equações de Maxwell, que unificaram eletricidade e magnetismo, foram formuladas em linguagem matemática que revelou simetrias não óbvias. A geometria riemanniana, desenvolvida por razões puramente matemáticas, tornou-se base da teoria da relatividade geral de Einstein.

A "eficácia irrazoável" de Wigner (1960) pode ser analisada em dois subproblemas distintos que têm respostas potencialmente diferentes. O primeiro é por que a matemática já existente se aplica tão bem à física, o problema da aplicação. O segundo é por que a matemática desenvolvida sem motivação física acaba, frequentemente, encontrando aplicações físicas décadas depois, o problema da pré-adaptação. A resposta naturalista da co-evolução aborda principalmente o primeiro problema; o segundo permanece mais misterioso e pode favorecer uma posição mais próxima do platonismo estrutural (Wigner, 1960, p. 1-14).

Como explicar isso? Os platonistas têm uma resposta. A matemática descreve a estrutura da realidade. O mundo físico exemplifica estruturas matemáticas abstratas, então não é surpreendente que a matemática funcione. Os nominalistas têm mais dificuldade, pois se a matemática é apenas convenção ou jogo formal, por que o mundo físico obedeceria a ela? É um mistério.

Uma resposta naturalista é, conforme aludimos anteriormente, que a matemática e a física co-evoluíram. A matemática foi desenvolvida, em grande parte, para resolver problemas práticos e teóricos, inspirados por fenômenos físicos. Assim, as teorias matemáticas que sobreviveram são aquelas que se mostraram úteis. Portanto, não é milagroso que a matemática se aplique ao mundo; ela foi moldada por essa aplicação.

Mas isso não explica completamente o nosso questionamento. Muitas teorias matemáticas foram desenvolvidas séculos antes de encontrarem suas aplicações, como a teoria dos números na criptografia moderna, ou a geometria não-euclidiana na relatividade. Parece haver uma sintonia fina entre matemática pura e física que desafia explicações puramente históricas.

O naturalismo matemático modesto pode acomodar esse fato. Se a matemática é contínua com a ciência empírica, então não há separação radical entre conhecimento matemático e conhecimento físico. Ambos são parte da mesma empresa de entender o mundo. A matemática fornece estruturas conceituais

que tornam possível a física. A física fornece problemas que estimulam o desenvolvimento da matemática. A relação é de interdependência, não de milagre.

4. Ética da Matemática

Pode parecer que a ética seja irrelevante para a matemática. A matemática é pura, abstrata, desconectada de preocupações mundanas sobre certo e errado. Matemáticos provam teoremas, não fazem julgamentos morais. Mas essa visão é ingênua. A matemática tem consequências, e consequências têm dimensões éticas.

A ilusão de neutralidade ética da matemática tem raízes na própria concepção platônica de que se os objetos matemáticos existem num reino eterno e imutável, separado do mundo humano, então a matemática pura não poderia ser moralmente carregada. Mas mesmo que se aceite alguma versão do platonismo, isso não isenta a prática matemática, a atividade humana de descobrir, comunicar e aplicar matemática, de considerações éticas. A distinção relevante é entre a matemática como estrutura abstrata e a matemática como prática social, e é sobre esta última que as questões éticas incidem.

G.H. Hardy (1940/1992), em *A Mathematician's Apology*, declarou orgulhosamente que seu trabalho em teoria dos números era perfeitamente inútil, sem aplicações práticas. Isso era uma virtude em sua visão. A matemática pura não se contamina com a utilidade. Mas essa posição é insustentável agora. A teoria dos números, que Hardy (1940/1992) considerava inútil, é fundamental para a criptografia moderna. Cada vez que você usa internet banking, depende de teoremas sobre números primos.

O caso de Hardy (1940/1992) é instrutivo porque revela que a separação entre matemática pura e aplicada não é apenas epistemológica, mas também ética e política; ela funcionou historicamente como uma forma de imunidade moral para os matemáticos puros, que podiam cultivar sua disciplina sem se preocupar com consequências. Com a ubiquidade das aplicações tecnológicas da matemática pura, essa imunidade deixou de ser disponível. O matemático contemporâneo que trabalha em teoria dos números, álgebra abstrata ou análise funcional não pode mais invocar a “inutilidade” de seu trabalho como escudo moral (Hardy, 1940/1992, p. 140-149).

Isso levanta a questão da responsabilidade. Se um matemático descobre um teorema que possibilita uma tecnologia perigosa, ele é responsável pelas consequências? John von Neumann trabalhou no Projeto Manhattan, desenvolvendo a bomba atômica. A matemática que ele desenvolveu foi essencial. Ele foi parcialmente responsável por Hiroshima? Ou a responsabilidade é apenas de quem decidiu usar a bomba? Isso nos faz pensar, de modo geral, na natureza do conhecimento que trazemos ao mundo e divulgamos, e da nossa responsabilidade ao fazer isso.

A questão da responsabilidade dos cientistas pelo uso de seu trabalho pode ser enquadrada a partir de distinções da teoria da responsabilidade moral. Uma distinção relevante é entre responsabilidade causal e responsabilidade moral. Von Neumann foi causalmente responsável pelos desenvolvimentos matemáticos que tornaram a bomba possível, mas a responsabilidade moral pela decisão de usá-la recai sobre outros. Outra distinção relevante é entre responsabilidade prospectiva (obrigação de antecipar e mitigar consequências) e responsabilidade retrospectiva (prestação de contas por consequências passadas). Os matemáticos têm, certamente, responsabilidade prospectiva de considerar as aplicações previsíveis de seu trabalho, mesmo que a responsabilidade retrospectiva por usos imprevistos seja mais controversa. E isso nos faz pensar que talvez ele tenha uma certa responsabilidade moral pelo uso. O assunto é espinhento.

Uma questão análoga surge com a criptografia. Matemáticos desenvolvem algoritmos criptográficos que protegem a privacidade e a segurança. Mas os mesmos algoritmos podem proteger criminosos e terroristas. Governos querem *backdoors* na criptografia; mas os criptógrafos argumentam que *backdoors* comprometeriam a segurança para todos. Isso não é uma disputa puramente técnica, mas tem dimensões éticas sobre privacidade, segurança e autonomia.

O debate sobre *backdoors* na criptografia ilustra um conflito entre valores que a ética filosófica tem recursos para analisar. Do ponto de vista utilitarista, a questão é empírica. Os benefícios de segurança proporcionados pelos *backdoors* superam os custos em privacidade e vulnerabilidade? A maioria dos criptógrafos argumenta que não, pois um backdoor acessível a governos democráticos é também explorável por governos autoritários e por atores maliciosos. Do ponto de vista deontológico, a questão é sobre direitos. Há um direito à comuni-

cação privada que não pode ser violado mesmo por razões de segurança pública? A análise filosófica não resolve a disputa, mas clarifica o que está em jogo.

A inteligência artificial é outro caso. *Machine learning* depende profundamente de matemática, como álgebra linear, cálculo, estatística, otimização. Algoritmos de IA estão cada vez mais poderosos e opacos. Tomam decisões que afetam vidas sem que possamos nem explicar por quê. Os matemáticos que desenvolvem esses algoritmos têm responsabilidade pelas consequências discriminatórias?

E o problema da opacidade dos algoritmos de aprendizado profundo tem uma dimensão matemática específica, a saber, a alta dimensionalidade dos espaços de parâmetros torna a interpretabilidade analítica praticamente impossível. Isso gera o que a literatura chama de "problema da caixa-preta", de o sistema produzir resultados corretos sem que possamos reconstruir o raciocínio que os gerou. Do ponto de vista ético, isso viola o que alguns teóricos chamam de princípio da explicabilidade, de que decisões que afetam significativamente a vida das pessoas devem ser acompanhadas de uma explicação que permita contestá-las. Matemáticos e engenheiros que desenvolvem sistemas que violam esse princípio têm responsabilidade de buscar alternativas mais interpretáveis, mesmo que menos eficientes.

Alguns argumentam que os matemáticos não têm uma responsabilidade especial. A matemática é uma ferramenta, e ferramentas podem ser usadas para o bem ou para o mal. Não culpamos as ferramentas nem seus criadores pelo mau uso. Um martelo pode construir uma casa ou matar alguém. Não culpamos o inventor do martelo por assassinatos. Similarmente, a matemática pode ser usada destrutivamente, mas isso não torna os matemáticos responsáveis.

Essa analogia é enganosa. Matemáticos não são como criadores de ferramentas que fazem produtos neutros. A matemática já vem carregada de valores e suposições. A escolha de quais problemas trabalhar, que métodos usar, que aplicações buscar, todas envolvem julgamentos de valor. Além disso, matemáticos frequentemente trabalham com aplicações específicas em mente. Um criptógrafo militar não é inocente sobre o uso de seu trabalho.

Há também questões sobre acesso e equidade. A matemática avançada é dominada por poucas instituições de elite em países ricos. Isso cria desigualda-

des. Talentos matemáticos de países pobres têm menos oportunidades. Mulheres e minorias são sub-representadas na matemática. Isso não é apenas injusto individualmente, mas também desperdiça talento socialmente e distorce quais problemas são estudados.

A sub-representação de mulheres e minorias na matemática não é um fenômeno natural, mas o resultado de barreiras sociais e institucionais documentadas, como estereótipos de aptidão, ausência de modelos, ambientes hostis e processos de avaliação enviesados. Do ponto de vista filosófico, isso constitui uma forma de injustiça epistêmica estrutural, pois as perspectivas e os problemas considerados importantes pela comunidade matemática refletem a composição demográfica dessa comunidade, o que significa que certas linhas de investigação são sistematicamente sub-exploradas. A diversificação da comunidade matemática não é apenas uma questão de equidade individual; é uma questão de qualidade epistêmica coletiva.

A comunidade matemática poderia fazer mais para promover a diversidade e a inclusão. Poderia tornar a matemática mais acessível através de uma melhor educação, de recursos abertos, de mentoria. E reconhecer vieses em quais problemas são considerados importantes ou elegantes. A teoria dos números é vista como nobre, a matemática aplicada como inferior. Essa hierarquia reflete valores, não fatos sobre a natureza da matemática.

Há também questões sobre publicação e crédito. Competição por crédito pode encorajar segredo, conflitos, comportamento antiético. Casos de plágio, má conduta, disputas de prioridade ocorrem constantemente na matemática. A pressão para publicar pode comprometer a qualidade, além da revisão por pares poder ser tendenciosa. Essas não são apenas questões sociológicas, mas têm dimensões éticas sobre a integridade e o bem comum.

Finalmente, há questões sobre como a matemática é ensinada. A educação matemática frequentemente aliena estudantes, fazendo-os sentir que a matemática é para gênios, não para pessoas normais. Isso desperdiça o talento e perpetua as desigualdades. Matemáticos têm a responsabilidade de tornar a matemática acessível, de mostrar sua beleza e poder, não apenas sua dificuldade. Além disso, numa sociedade cada vez mais tecnológica, a compreensão matemática básica é essencial para a cidadania. Decisões sobre políticas públicas frequentemente envolvem conceitos matemáticos e estatísticos. Cidadãos que

não entendem esses conceitos são excluídos do debate democrático. Matemáticos têm a responsabilidade de contribuir para educação pública e divulgação científica.

A questão da literacia matemática como condição de participação democrática conecta a ética da matemática à filosofia política. Numa democracia deliberativa, a qualidade da deliberação pública depende da capacidade dos cidadãos de avaliar argumentos quantitativos, estatísticas de saúde pública, modelos econômicos, projeções climáticas. Quando essa capacidade é distribuída desigualmente, o poder decisório efetivo concentra-se nas mãos de quem possui formação técnica, independentemente dos procedimentos formais de votação. A responsabilidade dos matemáticos pela educação pública não é, portanto, apenas caridosa; é uma exigência da democracia.

5. Considerações Finais

Examinei três problemas filosóficos centrais sobre a matemática, a saber, a natureza dos objetos matemáticos, o conhecimento das verdades matemáticas e as responsabilidades éticas dos matemáticos. Em cada caso, mostrei que há tensões sem resoluções fáceis, oposições que persistem. Minha posição através desses debates é pragmática e pluralista. Na metafísica, propus o naturalismo modesto de aceitar que a matemática postula entidades abstratas, mas reconhecer que isso é um mistério ontológico que não sabemos resolver. Não finjo que números claramente existem nem que claramente não existem. Ambas as posições têm problemas. Vivemos com o mistério, com o que, pragmaticamente, dada a utilidade da matemática nas ciências, vale a pena conviver.

Esse naturalismo modesto não é uma capitulação agnóstica, mas uma posição positiva; ele afirma que a ontologia matemática é um problema em aberto que não impede a prática matemática nem a avaliação de suas implicações éticas. A filosofia da matemática não precisa esperar a resolução da querela platonismo/nominalismo para se pronunciar sobre as responsabilidades dos matemáticos ou sobre a distribuição social do conhecimento matemático. A separação entre metafísica e ética, nesse domínio como em outros, é metodológica, não substantiva.

Na epistemologia, propus um certo pluralismo, no qual diferentes tipos de conhecimento matemático podem ter diferentes status epistemológicos. A aritmética básica talvez seja *a priori*. A geometria talvez seja empírica. A lógica talvez seja convencional. Não há por que forçar toda a matemática em uma categoria única. E talvez a distinção *a priori* x *a posteriori* seja menos clara do que tradicionalmente se pensou.

Esse pluralismo epistemológico tem uma implicação metodológica importante, a de que a avaliação de afirmações matemáticas específicas deve ser feita caso a caso, atendendo às características particulares de cada domínio, em vez de aplicar um critério uniforme derivado de uma teoria geral do conhecimento matemático. Isso é consonante com a prática matemática real, em que diferentes campos, teoria dos números, topologia, estatística, lógica, têm metodologias e padrões de justificação parcialmente distintos.

Na ética, propus que os matemáticos têm responsabilidades como qualquer cidadão, mas também responsabilidades especiais dado o conhecimento especializado e o poder que a matemática confere. Assim, os matemáticos não podem fingir neutralidade em seu trabalho, mas devem considerar as consequências de seu trabalho, promover acesso equitativo e educar responsavelmente.

A responsabilidade ética dos matemáticos é, portanto, tripartite, dado que é a responsabilidade prospectiva em relação às aplicações do seu trabalho, a responsabilidade distributiva em relação ao acesso equitativo ao conhecimento matemático, e a responsabilidade cívica em relação à literacia matemática da população. Nenhuma dessas responsabilidades é absoluta, todas são ponderadas por outros valores e circunstâncias, mas nenhuma pode ser simplesmente descartada por apelo à pureza ou à neutralidade da matemática.

Essas posições não eliminam tensões. As tensões persistem. Mas oferecem uma maneira de conviver produtivamente com elas, reconhecendo a complexidade ao invés de forçar simplificações que distorcem a realidade. A matemática é um fenômeno multifacetado, simultaneamente abstrato e aplicado, *a priori* e conectado à experiência, puro e eticamente carregado. Uma filosofia que não reconhece todas essas dimensões é inadequada.

Este capítulo prepara o terreno para os próximos. Num capítulo posterior, sobre Filosofia da Ciência da Informação, enfrentaremos algumas questões relacionadas. A informação, como os números, parece ser abstrata. Um bit não

é um objeto físico, mas um padrão que pode ser realizado em diferentes substratos. Dados, estruturas de dados, algoritmos, todos são abstratos de certa forma. Mas, diferentemente dos números, a informação está profundamente conectada a processos físicos. Ela não pode existir sem um substrato material. Isso criará tensões metafísicas diferentes.

A transição da matemática para a ciência da informação é, ela própria, filosoficamente reveladora. A matemática lida com estruturas abstratas que podem ser instanciadas em qualquer substrato, ou em nenhum. A informação, por contraste, é sempre informação em algum meio, para algum receptor, em algum contexto. Essa diferença, entre o abstrato puro e o abstrato encarnado, é o que torna a filosofia da informação um campo distinto, com seus próprios problemas metafísicos, epistemológicos e éticos.

Além disso, a ciência da informação enfrentará questões epistemológicas sobre como conhecemos através da informação mediada pela tecnologia. E questões éticas sobre privacidade, acesso e poder informacional. A matemática preparou o terreno para essas discussões, fornecendo exemplos de objetos abstratos e de conhecimento *a priori.* Mas a informação tem características distintivas que exigirão uma análise própria.

A tensão entre abstrato e concreto, universal e particular, puro e aplicado, perpassa múltiplas disciplinas neste livro. Vimos, na antropologia, com a cultura; na linguística, com as línguas; na matemática, com os números... E veremos repetidamente. É a tensão constitutiva de tentar entender fenômenos que transcendem os objetos físicos particulares, mas estão realizados neles. Não há fórmula mágica para resolver a tensão. Há apenas o trabalho paciente de articular problemas, avaliar posições, fazer julgamentos informados.

Da abstração matemática, passamos agora à informação, aquilo que pode ser registrado, armazenado, recuperado, transmitido. A informação é, de certa forma, a matemática tornada comunicável, a estrutura abstrata encarnada em suportes materiais. E com isso entramos no coração do mundo digital.

5

Filosofia da Ciência da Informação

1. Por que uma Filosofia da Ciência da Informação?

Quando um cientista da informação afirma que "a informação flui através de redes", quando um arquivista declara que "este documento contém informação sobre eventos históricos", quando um bibliotecário diz que "livros são portadores de informação", todos estão fazendo afirmações que pressupõem respostas a questões filosóficas fundamentais. O que é informação? Que tipo de entidade é? Onde existe? No mundo físico, em mentes, em ambos, ou em nenhum? Como se relaciona com os objetos físicos que supostamente a contêm ou carregam? Que responsabilidades éticas temos ao criar, organizar, preservar e disseminar informação? Essas questões não podem ser respondidas por métodos empíricos, estatísticas de uso, ou análises técnicas de sistemas. Elas exigem argumentação e fundamentação filosófica explícita.

A afirmação de que as práticas informacionais pressupõem respostas a questões filosóficas pode ser articulada com o conceito wittgensteiniano de "gramática profunda" (*Tiefengrammatik*), segundo o qual toda forma de vida, inclusive a prática profissional do arquivista, do bibliotecário, do cientista da informação, opera com uma lógica implícita que antecede o método empírico e que só a análise filosófica pode tornar visível.

Este capítulo dá continuidade à trilha que estamos percorrendo, e que se iniciou com a antropologia, a linguística e a matemática. A antropologia nos mostrou que a cultura é o meio pelo qual os humanos criam significado; a linguística, que a linguagem é o veículo pelo qual esse significado é transmitido; e, finalmente, a matemática, nos mostrou a estrutura abstrata que subjaz à própria possibilidade de representação. Agora, a ciência da informação nos mostra o que acontece quando a linguagem se torna registro, quando a fala, que nor-

malmente se esvai em meio ao ar, se transforma em algo que pode ser armazenado, recuperado, transmitido através do tempo e do espaço. A informação é, de certa forma, a linguagem tornada durável. E com essa durabilidade vêm novas questões. O que acontece com o significado quando ele é separado de seu contexto original de enunciação? Como garantir que a informação preservada seja autêntica? Que responsabilidades temos para com aqueles que produzirão informação no futuro, a partir dos registros que deixamos?

A caracterização da informação como "linguagem tornada durável" pode ser aprofundada com a distinção de Paul Ricœur (1976) entre o discurso falado e o discurso escrito. Na escrita, o significado se "distancia" da intenção do autor e da situação original de enunciação, ganhando uma autonomia semântica que é precisamente o que torna possível, e necessária, a interpretação. A ciência da informação pode ser lida como a disciplina que administra institucionalmente as consequências dessa autonomia semântica.

A filosofia da ciência da informação investiga os fundamentos conceituais, metafísicos, epistemológicos e éticos do estudo científico da informação. Não se trata de uma subdisciplina da ciência da computação, da comunicação ou da gestão, embora dialogue com todas essas áreas. Trata-se de identificar e investigar sistematicamente os problemas conceituais que emergem quando examinamos criticamente o que fazemos quando criamos, organizamos, preservamos, recuperamos e disseminamos a informação.

Considere três tipos de perguntas sobre a informação. Perguntas empíricas são respondidas pela ciência, por exemplo, quantos usuários acessaram o repositório digital no último mês, ou qual a taxa de deterioração de documentos em papel em condições específicas de temperatura e umidade, ou ainda, que padrões de navegação os usuários exibem em bibliotecas digitais. Perguntas formais são respondidas pela matemática, lógica ou ciência da computação. Qual algoritmo de indexação maximiza precisão e revocação? Como formalizar relações semânticas em uma ontologia? Qual a complexidade computacional de um sistema de recuperação de informação? Finalmente, os problemas filosóficos são solucionados pela argumentação conceitual. Assim, saber o que distingue a informação de mero dado, ou se um documento digital é o mesmo objeto que sua contraparte física, ou se arquivistas têm o direito de destruir documen-

tos, ou uma responsabilidade absoluta de preservar tudo, essas são questões intimamente filosóficas.

A tripartição em perguntas empíricas, formais e filosóficas corresponde à distinção clássica entre questões de fato, questões formais e questões de valor/conceito. No entanto, a distinção não é estanque; perguntas aparentemente empíricas (como "que padrões de busca os usuários exibem?") pressupõem conceitos filosóficos carregados de relevância, necessidade e intenção; e perguntas filosóficas (como "o que distingue dado de informação?") têm implicações diretas sobre o design de sistemas técnicos.

É este conjunto de perguntas que constitui o domínio próprio da filosofia da ciência da informação. Enquanto as ciências da informação investigam como os sistemas funcionam, como os usuários se comportam, como as tecnologias podem ser otimizadas, a filosofia investiga o que são esses sistemas, se e por que importam, quais normas devem governá-los, e que tipos de valores expressam ou realizam.

Alguém poderia objetar que a ciência da informação é uma disciplina prática, orientada a problemas concretos de organização, preservação e acesso à informação, de modo que a reflexão filosófica abstrata seria um luxo desnecessário. Mas essa objeção pressupõe falsamente que práticas informacionais são teoricamente neutras, quando na verdade toda prática incorpora pressupostos teóricos e filosóficos que podem e devem ser examinados, e ignora que muitos problemas enfrentados por profissionais da informação são fundamentalmente filosóficos, disfarçados de questões técnicas. Isso certamente subestima o potencial da reflexão filosófica para iluminar, questionar e aprimorar as práticas profissionais (Floridi, 2011).

Não existe prática teoricamente neutra. Quando um arquivista aplica o princípio da proveniência, está operacionalizando uma teoria sobre a identidade e a integridade documental que possui pressupostos ontológicos e epistemológicos (Duranti, 1998). Quando um bibliotecário utiliza a Classificação Decimal de Dewey, está aplicando uma epistemologia particular sobre a estrutura do conhecimento humano, uma epistemologia que privilegia certas áreas, marginaliza outras, e reflete valores culturais específicos (Shera, 1970). Quando um cientista da informação projeta um sistema de recomendação algorítmica, está implementando uma teoria sobre relevância, sobre o que significa satisfazer necessi-

dades informacionais, e sobre a relação entre sistema e usuário. Estes pressupostos filosóficos raramente são explicitados ou examinados criticamente na formação profissional. O resultado é que profissionais frequentemente operam com teorias implícitas, não articuladas, potencialmente contraditórias, que nunca foram submetidas a escrutínio filosófico.

Muitos problemas que aparecem como questões técnicas ou administrativas são, em sua essência, filosóficos. Na preservação digital, arquivistas enfrentam o problema de garantir a autenticidade de documentos em ambientes onde a modificação é trivial e imperceptível. Este não é apenas um problema técnico sobre assinaturas digitais, mas uma questão metafísica sobre a identidade através de mudança. O que torna um documento digital em 2026 o mesmo documento que foi criado em 2006, quando nenhum bit físico permanece idêntico, e o formato e o contexto tecnológico são radicalmente diferentes? Esta é a versão informacional do antigo problema do navio de Teseu (Thibodeau, 2002). O problema do navio de Teseu aplicado à preservação digital tem uma especificidade que o distingue da versão clássica. No caso dos objetos físicos, a questão é sobre a substituição gradual das partes materiais; no caso dos documentos digitais, a questão é sobre a relação entre instâncias, formatos e substratos radicalmente heterogêneos que nenhuma continuidade material conecta. A identidade do documento digital é uma identidade funcional e institucional, não física.

Na seleção de acervo, os bibliotecários confrontam regularmente demandas conflitantes sobre que materiais incluir ou excluir de coleções. Alguns defendem que as bibliotecas devem ser absolutamente neutras, adquirindo materiais de todos os pontos de vista. Outros argumentam que certos materiais, como literatura que promove ódio ou desinformação científica, não merecem espaço em coleções. Este não é apenas um problema de política de desenvolvimento de coleções, mas um dilema ético profundo sobre liberdade intelectual, paternalismo, e responsabilidade social (Mill, 1859/1991).

Na tensão entre acesso e privacidade em arquivos, os arquivistas custodiam documentos que contêm informações sensíveis sobre os indivíduos, e têm de decidir entre as demandas contraditórias dos pesquisadores que demandam acesso para fins legítimos e os descendentes que exigem proteção da privacidade de seus antepassados. Este não é meramente um problema de interpretar

regulamentos, mas um conflito filosófico fundamental entre os valores de transparência histórica e dignidade individual, o direito de saber contra o direito de ser esquecido (Derrida, 2001).

No viés algorítmico, sistemas de IA usados para recuperação de informação, moderação de conteúdo, ou recomendação reproduzem e amplificam vieses presentes nos dados de treinamento. Este não é apenas um problema técnico de engenharia de algoritmos, mas uma questão epistemológica sobre a objetividade, e uma questão ética e política sobre justiça e poder (O'Neil, 2016; Pasquale, 2015).

Esses exemplos ilustram que problemas aparentemente técnicos revelam-se, sob exame, filosóficos. Soluções técnicas podem ser implementadas apenas depois que questões filosóficas subjacentes são enfrentadas. A reflexão filosófica pode e deve informar, questionar e aprimorar as práticas profissionais. Quando os profissionais da informação entendem os fundamentos filosóficos de suas práticas, tornam-se mais capazes de justificar decisões, antecipar problemas, criticar normas estabelecidas, navegar dilemas e comunicar-se com outras disciplinas.

2. Metafísica da Informação

A metafísica investiga a natureza fundamental da realidade. Que tipos de coisas existem, quais são suas propriedades essenciais, como se relacionam umas com as outras? Aplicada à informação, a metafísica pergunta o que é a informação. Que tipo de entidade é? Existe independentemente de mentes? Como se relaciona com os objetos físicos que a carregam?

A pergunta sobre o que é informação é enganosamente simples. Todos usamos o termo constantemente, parecendo saber perfeitamente bem o que significa. Mas quando tentamos definir “informação” precisamente, descobrimos que o conceito é notavelmente elusivo e multifacetado. Diferentes teorias e diferentes disciplinas adotam concepções radicalmente diferentes, frequentemente incompatíveis.

Buckland (1991), em seu influente artigo *Information as Thing*, distingue três sentidos principais de informação. Informação-como-processo é o ato de informar, de comunicar o conhecimento, que modifica o que o receptor sabe. Nesse

sentido, a informação é o evento, e não a entidade. Informação-como-conhecimento seria aquilo que é comunicado, o conhecimento transmitido. Esse sentido é próximo de nossa fala ordinária quando dizemos que adquirimos muita informação sobre um tópico. Informação-como-conhecimento é intangível, mental, subjetiva, existe em mentes, não no mundo físico. Informação-como-coisa são objetos físicos que são considerados informativos, como documentos, dados, objetos. Um livro, um arquivo digital, um espécime de museu podem todos ser informação neste sentido. Informação-como-coisa é tangível, física, objetiva, pode ser manipulada, armazenada, transmitida fisicamente.

A tripartição de Buckland (1991) tem antecedentes na distinção peirceana entre tipo (*type*) e ocorrência (*token*), pois a informação-como-coisa é uma ocorrência particular; a informação-como-conhecimento aproxima-se do tipo ou do conteúdo proposicional. A contribuição específica de Buckland (1991) é mostrar que os sistemas de informação operam primariamente no nível das ocorrências materiais, enquanto o interesse epistêmico reside nos tipos semânticos, tensão que nenhum sistema de informação resolve completamente.

Buckland (1991) argumenta que apenas informação-como-coisa pode ser gerenciada por sistemas de informação, pois apenas coisas físicas podem ser armazenadas, recuperadas e transmitidas por meios técnicos. O conhecimento em mentes não pode ser diretamente acessado ou transferido, e processos de informar são eventos efêmeros que ocorrem e terminam. Sistemas de informação operam sobre objetos físicos que representam ou incorporam a informação.

Esta análise tem implicações práticas importantes. Se Buckland (1991) estiver certo, então a ciência da informação deveria focar em objetos informacionais físicos (documentos, dados, arquivos), e não em estados mentais ou processos comunicativos. A disciplina seria fundamentalmente sobre gestão de coisas, não de conhecimento diretamente. Mas esta conclusão é controversa. Muitos teóricos argumentam que reduzir informação a objetos físicos perde exatamente o que torna a informação interessante, que é o seu conteúdo semântico, seu significado, sua capacidade de representar aspectos do mundo e afetar a cognição. Um livro cujo conteúdo ninguém compreende, escrito em língua morta completamente esquecida, é ainda informação? Parece que não, pois perdeu sua capacidade de informar. Mas como objeto físico, nada mudou. Logo, a informação não pode ser simplesmente identificada com os objetos físicos.

Outra distinção crucial é entre informação semântica, com significado, e informação sintática, sem significado. Esta distinção emerge claramente na teoria matemática da comunicação de Shannon e Weaver (1949). Shannon (1949) desenvolveu uma teoria matemática rigorosa para quantificar a informação transmitida através de canais de comunicação. Sua medida de informação, baseada em entropia, é puramente sintática, pois depende apenas de probabilidades de símbolos, não de seus significados. Uma sequência aleatória de letras contém máxima informação no sentido de Shannon (1949) (alta entropia, baixa previsibilidade), mas nenhum significado. Uma frase significativa contém menos informação no sentido de Shannon (1949) (menor entropia devido a correlações linguísticas), mas é o que intuitivamente chamamos informativo.

Esta divergência entre informação sintática e informação semântica gera uma confusão persistente. Shannon (1949) explicitamente advertiu que sua teoria não aborda o significado, afirmando que aspectos semânticos de comunicação são irrelevantes para o problema de engenharia. Sua teoria mede a quantidade de informação transmitida, não sua qualidade ou significado. Mas cientistas da informação, bibliotecários e arquivistas estão primariamente interessados em informação semântica. Quando catalogamos um livro ou indexamos um documento, estamos preocupados com seu conteúdo significativo, não com o número de bits necessários para codificá-lo. Quando usuários buscam informação sobre mudanças climáticas, querem significado, não sequências sintáticas. Isto levanta uma questão filosófica profunda sobre se a informação semântica é um tipo radicalmente diferente da informação sintática, ou se há uma relação sistemática entre elas. Pode-se reduzir a semântica à sintaxe, ou a semântica é irredutível?

A irredutibilidade da semântica à sintaxe é defendida por John Searle (1980) no argumento do "quarto chinês", no qual um sistema pode processar símbolos sintaticamente corretos sem qualquer compreensão semântica. Se Searle (1980) está certo, nenhum sistema de recuperação de informação, por mais sofisticado, acede genuinamente ao significado, ele manipula sintaxe que, para usuários humanos, tem significado. Isso tem implicações diretas sobre os limites epistêmicos dos sistemas de recuperação.

Fred Dretske (1981), em *Knowledge and the Flow of Information*, propõe uma teoria que tenta conectar informação sintática e semântica. Segundo

Dretske (1981), informação semântica é informação que um sistema recebe e que causa crenças verdadeiras naquele sistema. Um sinal carrega informação semântica que p, onde p é proposição, se e somente se a presença do sinal indica que p é verdadeira. Esta análise tenta naturalizar significado, explicá-lo em termos de relações causais e informacionais que não pressupõem significado. Os críticos argumentam que a teoria de Dretske (1981) não captura plenamente a semântica porque a informação pode estar presente causalmente, sem ser reconhecida como significativa. Anéis de árvore carregam a informação sobre a idade da árvore, mas só têm significado para organismos capazes de interpretá-los. Isto sugere que a semântica requer interpretadores, não emerge automaticamente das relações informacionais.

Uma tensão perpassa as discussões sobre informação, a saber, a informação parece ser simultaneamente física e mental, objetiva e subjetiva. Por um lado, a informação está inscrita em substratos físicos, como padrões em papel, estados magnéticos em discos, configurações de neurônios, sendo transmitida causalmente através de processos físicos. Por outro lado, a informação tem um conteúdo representacional, refere-se a coisas além de si mesma, e este conteúdo parece depender de mentes que a interpretam. Esta dualidade reflete a natureza genuinamente dual da informação.

Considere um documento histórico, digamos uma carta escrita por Dom Pedro II. Fisicamente, é um pedaço de papel com marcas de tinta, moléculas organizadas em padrões específicos. Estas propriedades físicas podem ser completamente descritas por física e química sem referência a significado. Mas o documento também tem conteúdo semântico, pois expressa pensamentos de Dom Pedro, refere-se a eventos históricos, pode ser verdadeiro ou falso. Estas propriedades semânticas não são redutíveis a propriedades físicas.

Os filósofos da mente debatem intensamente como as propriedades mentais relacionam-se com as propriedades físicas, e se dividem, grosso modo, em três posições principais. O dualismo afirma que propriedades mentais e físicas são fundamentalmente distintas e irredutíveis. O fisicalismo reducionista nos diria que propriedades mentais são idênticas a propriedades físicas, ou redutíveis a elas. Versões extremas do fisicalismo implicam que a informação semântica não existe realmente; apenas informação sintática, física, existe. Por sua vez, o fisicalismo não-reducionista sustenta que propriedades mentais sobrevêm

a propriedades físicas, sem serem idênticas ou redutíveis a elas. A superveniência do mental com relação ao físico significa que não pode haver diferença mental sem alguma diferença física correspondente, sem que isso implique a redução de um domínio a outro, mas meramente a dependência.

Aplicado à informação, o fisicalismo não-reducionista sugeriria que a informação semântica sobrevém ao substrato físico, mas possui uma autonomia conceitual. Um documento é constituído fisicamente por papel e tinta, mas seu conteúdo informacional não é idêntico a nem redutível a essas propriedades físicas. Mudanças no conteúdo informacional sempre envolvem mudanças físicas, mas múltiplas realizações físicas diferentes (papel, tela digital, áudio) podem carregar o mesmo conteúdo informacional. Esta posição é atraente para as ciências da informação, porque reconhece a base física dos sistemas informacionais, enquanto preserva a autonomia do nível semântico, permitindo que os profissionais da informação se concentrem em conteúdo e significado, sem negar que a informação requer substratos físicos.

A múltipla realizabilidade do conteúdo informacional, o mesmo conteúdo em substratos físicos distintos, é o argumento central do funcionalismo em filosofia da mente. Aplicado à informação, ela fundamenta a possibilidade de migração de formatos na preservação digital sem perda de identidade, pois o que se preserva é a função (o conteúdo), não o substrato. Isso, contudo, pressupõe que se possa especificar o conteúdo independentemente do substrato, tarefa que os metadados de preservação tentam realizar.

Luciano Floridi (2011) desenvolveu uma teoria sistemática e ambiciosa que tenta unificar diversos sentidos de informação. Em *The Philosophy of Information* (2011), Floridi (2011) propõe uma definição geral, de que as informações são dados bem formados e significativos. Esta definição possui três componentes. Dados são diferenças, variações, padrões. No nível mais fundamental, a informação consistiria de distinções, como presença versus ausência, zero versus um, este símbolo versus aquele. Sem diferença, não há informação. “Bem formados” significa que dados devem ter estrutura, seguir regras sintáticas. Sequências aleatórias de símbolos não constituem informação, se violam regras de formação. E “significativos” significa que dados bem formados devem ter semântica, referir-se a algo, representar estados de coisas. Esta é a dimensão que distingue a informação do mero dado sintático.

Floridi (2011) distingue ainda tipos de informação semântica. A informação factual representa estados de coisas que obtém no mundo; a informação instrucional não representa fatos mas guia a ação, como receitas, algoritmos, comandos; e a informação ficcional representa estados de coisas que não obtêm mas são internamente consistentes, como romances. Esta taxonomia é útil para ciências da informação porque diferentes tipos de informação exigem tratamentos diferentes. A informação factual em documentos de arquivo deve ser verificável e autêntica. A informação instrucional em manuais deve ser clara e executável. E a informação ficcional em literatura deve ser catalogada diferentemente de não-ficção.

Críticos de Floridi (2011) argumentam que sua definição é simultaneamente muito estreita e muito ampla. Muito estreita porque exclui informação que não é semântica, como a teoria de Shannon (1949). Muito ampla porque inclui tudo que é dado bem formado e significativo, tornando informação coextensivo com representação ou conteúdo. Além disso, exigir que a informação seja bem formada e significativa torna a informação dependente de convenções linguísticas e capacidades interpretativas. Isto implicaria que nada seria informação antes de existirem seres capazes de interpretação, o que parece contraintuitivo, pois certamente havia informação no universo antes de surgirem mentes.

A diversidade de teorias e definições de informação sugere que não há conceito único correto. Diferentes contextos e propósitos exigem diferentes conceitualizações. Este pluralismo conceitual reflete um fenômeno genuinamente multifacetado. Carl Bereiter (2002) argumenta explicitamente pelo pluralismo com relação à informação e ao conhecimento. Sua ideia é de que não há uma essência única de informação que todas as teorias tentam capturar. Ao invés, "informação" é termo que cobre um conjunto de fenômenos relacionados, que compartilham semelhanças de família, no sentido de Wittgenstein (1953/2009), mas não possuem um núcleo comum.

O recurso às "semelhanças de família" (*Familienähnlichkeiten*) de Wittgenstein (1953/2009) não é apenas uma estratégia de resolução de disputas terminológicas, pois implica que a tarefa do filósofo da ciência da informação não é descobrir a essência da informação, mas mapear os usos do conceito, suas conexões e seus contextos de aplicação. Isso aproxima a filosofia da informa-

ção da filosofia analítica da linguagem ordinária (Austin, 1962; Ryle, 1949) mais do que da metafísica essencialista.

Esta posição pragmática é metodologicamente saudável. Evita disputas verbais sobre a definição correta de informação, permitindo que diferentes teorias coexistam, cada uma útil para propósitos específicos. A teoria de Shannon (1949) é excelente para a engenharia de comunicação. A teoria semântica de Dretske (1981) é adequada para a epistemologia naturalizada. A análise de Buckland (1991) é útil para a gestão de sistemas de informação. A taxonomia de Floridi (2011) organiza um domínio conceitual complexo. Profissionais da informação podem adotar o pluralismo pragmático, usando conceitos de informação apropriados ao contexto, sem comprometimento ontológico forte com uma teoria única; ou podem se comprometer fortemente com uma das teorias, mas tendo em mente as implicações desses comprometimentos.

Seja o que for a informação, as questões sobre sua relação com a realidade física são centrais para a metafísica da ciência da informação. A informação é uma propriedade fundamental do universo, algo que existe desde sempre, independentemente de haver mentes para interpretá-la? Ou ela é um fenômeno derivado, que só emerge em contextos biológicos e sociais, quando criaturas como nós começam a interpretar padrões físicos como tendo significado? Duas posições metafísicas abrangentes disputam seriamente a resposta.

De um lado, temos o realismo informacional. Essa é a tese de que a informação existe objetivamente, no mundo, independentemente de haver mentes ou intérpretes para percebê-la. Padrões físicos e correlações causais, por exemplo, constituem informação por si só, mesmo que ninguém jamais os reconheça ou utilize. Os exemplos clássicos são fascinantes. Pense no DNA. A molécula de DNA carrega informação genética, i.e., uma sequência de nucleotídeos que "codifica" instruções para a síntese de proteínas. Essa informação existia muito antes de qualquer biólogo existir para decifrá-la, muito antes de sequer sabermos que ela estava lá. Os fósseis, da mesma forma, contêm informação sobre organismos que viveram há milhões de anos (sobre sua anatomia, sua idade, o ambiente em que viveram), mesmo quando estão soterrados e ninguém os encontrou ainda. Os anéis de crescimento das árvores registram informação sobre o clima histórico (anos de seca, anos de chuva abundante), e essa informação está ali, inscrita na madeira, quer alguém a leia ou não. Para o

realista informacional, esses são casos claros de informação que existe objetivamente, como uma propriedade do mundo físico, à espera de ser descoberta.

Do outro lado, temos o construtivismo informacional. Essa posição sustenta que a informação não é uma propriedade intrínseca dos objetos ou eventos, mas sim algo que as mentes atribuem ou interpretam a partir de padrões físicos. Não existe informação "ali fora" esperando para ser encontrada; o que existe são apenas padrões materiais. A informação surge apenas quando um sistema cognitivo (como um cérebro humano, ou talvez um sistema de IA) trata esses padrões como se representassem algo. A informação, nessa visão, é intrinsecamente relacional, pois só existe na relação entre um suporte físico (o padrão) e um intérprete (a mente que atribui significado). Para o construtivista, chamar o DNA de "informação" é uma projeção da nossa perspectiva humana sobre um processo puramente químico. O DNA, dizem eles, não "codifica" nada intrinsecamente; ele é uma molécula que interage causalmente com outras moléculas (enzimas, RNAs, proteínas) através de forças físico-químicas perfeitamente ordinárias. O fato de descrevermos isso como "informação que carrega instruções" é uma metáfora extremamente útil, sem dúvida, mas não é uma descrição literal de uma propriedade que a molécula possuiria em si mesma. A informação, nessa visão, está nos olhos de quem vê, não no mundo.

O construtivismo informacional aproxima-se da posição de Nelson Goodman (1978), pois o mundo não é dado, mas construído por meio de sistemas simbólicos. Para Goodman (1976), diferentes sistemas de símbolos, incluindo sistemas de classificação bibliográfica, ontologias, esquemas de metadados, não descrevem um mundo pré-dado, mas constituem mundos distintos. A implicação para a ciência da informação é que sistemas de organização do conhecimento são ontologicamente constitutivos, não meramente descritivos.

É tentador achar que esse debate é uma disputa abstrata e distante, coisa de filósofos trancados em suas torres de marfim. Mas não é. As implicações são profundamente práticas, especialmente para os profissionais da informação. Se o realismo estiver correto, então o trabalho de um arquivista, de um bibliotecário, de um cientista da informação é, em essência, um trabalho de descoberta. A informação já está lá, objetivamente estruturada no mundo ou nos documentos, e a tarefa do profissional é identificá-la, organizá-la e torná-la acessível. Os sistemas de classificação, como a Classificação Decimal de Dewey (CDD) ou a

Classificação Decimal Universal (CDU), seriam tentativas de refletir uma estrutura objetiva do conhecimento que existe independentemente de nós. Haveria uma maneira "certa" de classificar as coisas, e nosso trabalho seria nos aproximarmos dela. Se, por outro lado, o construtivismo estiver correto, então o trabalho do profissional da informação é um trabalho de construção. Não há uma ordem natural pré-existente a ser descoberta; o que há são decisões que tomamos, convenções que estabelecemos, formas de organizar que são mais ou menos úteis para determinados propósitos. Os sistemas de classificação não seriam espelhos da natureza, mas sim ferramentas pragmáticas, artefatos culturais que refletem nossos valores e interesses. A escolha entre uma visão e outra, portanto, muda radicalmente a autocompreensão do profissional e os critérios pelos quais avaliamos seu trabalho.

Há, no entanto, uma posição ainda mais radical. Alguns físicos e filósofos, indo além do realismo informacional, propõem que a informação não é apenas uma propriedade onipresente do universo, mas sim a propriedade fundamental da qual todas as outras derivam. O físico John Archibald Wheeler (1990), um dos gigantes da física do século XX, cunhou um slogan provocativo para essa visão, "*it from bit*". A expressão é enganosamente simples, mas carrega uma tese metafísica de tirar o fôlego. Wheeler (1990) está sugerindo que toda coisa (*every it*), cada partícula elementar, cada campo de força, cada constante da natureza, o próprio continuum espaço-temporal, deriva sua existência, sua função, seu próprio ser, de respostas a perguntas binárias, de escolhas do tipo "sim-ou-não", de bits de informação. Em outras palavras, a realidade física não é a base sobre a qual a informação repousa; é o contrário. A informação é a base, e a realidade física é uma consequência dela.

Essa interpretação, é preciso dizer, é motivada em parte por aspectos contraintuitivos da mecânica quântica. Antes de uma medição, um sistema quântico existe em uma superposição de estados, não possuindo valores definidos para suas propriedades. É o ato da medição (que podemos pensar como uma "pergunta" feita ao sistema) que "colapsa" a superposição e produz um resultado definido. Wheeler (1990) via nisso a sugestão profunda de que a realidade, no nível mais fundamental, não é feita de "coisas", mas de respostas a perguntas. O universo seria, nessa visão, um universo participativo em que a informação é constitutiva da própria realidade.

É preciso cautela. A tese de Wheeler (1990) é extremamente controversa, e poucos físicos a aceitam em sua formulação mais literal. Muitos a veem mais como uma metáfora poderosa do que como uma descrição literal do mundo. No entanto, versões mais moderadas da ideia de que a informação é fundamental têm ganhado cada vez mais tração na física teórica. A teoria quântica da informação, por exemplo, é um campo florescente que estuda os limites físicos do processamento de informação, tratando a informação como uma grandeza física fundamental, sujeita a leis tão básicas quanto as da termodinâmica. O princípio holográfico, uma conjectura ousada da física teórica, sugere que toda a informação contida em um volume de espaço pode ser representada em sua superfície bidimensional, o que implicaria que a informação, e não o espaço tridimensional, é a realidade mais fundamental. E desde os trabalhos pioneiros de Boltzmann e Maxwell no século XIX, os físicos reconhecem conexões profundas entre a entropia termodinâmica (uma medida da desordem de um sistema) e a informação. Mais recentemente, Rolf Landauer (1961) demonstrou que apagar um bit de informação tem um custo energético mínimo inevitável, um resultado que sugere que a informação não é uma abstração etérea, mas sim uma entidade com implicações físicas reais e mensuráveis.

Se a informação for, de fato, uma propriedade fundamental do universo, então a ciência da informação deixa de ser apenas uma disciplina aplicada, preocupada com a gestão de artefatos humanos (documentos, livros, arquivos digitais), e passa a ser um campo que investiga aspectos fundamentais da própria realidade. Seria uma reivindicação impressionante de status intelectual, colocando a ciência da informação ao lado da física e da cosmologia na busca pelos constituintes últimos do mundo. No entanto, é preciso ter cuidado antes de celebrar essa conclusão. Mesmo que os físicos usem o conceito de informação de maneira produtiva, não se segue automaticamente que a "informação" de que eles falam seja a mesma coisa que a "informação" com que lidam bibliotecários e arquivistas. A informação quântica, por exemplo, é puramente sintática; ela se refere a estados de sistemas físicos, a *qubits*, a probabilidades. Já a informação em documentos e bancos de dados é essencialmente semântica; ela depende de significado, de interpretação, de referência ao mundo. Não está nada claro se essas são duas instâncias de um mesmo fenômeno fundamental ou se são

dois conceitos que compartilham apenas o nome, como “manga” (a fruta) e “manga” (a parte da camisa). A questão permanece em aberto.

Há, ainda, uma questão metafísica específica que conecta informação e realidade de uma maneira ainda mais direta, e que tem implicações urgentes, por exemplo, para a biologia. Trata-se da relação entre informação e causalidade. A informação pode causar eventos físicos? Ou, de forma mais rigorosa, devemos dizer que apenas processos físicos causam eventos, e a “informação” é apenas uma descrição abstrata e conveniente desses processos?

Essa pergunta se torna urgentemente prática quando olhamos para a biologia molecular. Os biólogos falam o tempo todo em “informação genética”. Dizem, por exemplo, que um “gene para olhos azuis” *causa* a pessoa ter olhos azuis porque *transmite a informação* que instrui a síntese de proteínas que, por sua vez, afetam a pigmentação da íris. A informação, nessa narrativa, parece ser uma causa genuína, um agente que produz um efeito no mundo físico. Mas será que essa é uma maneira apropriada de falar? Os críticos argumentam que não. O DNA, dizem eles, não causa nada através da informação. O que acontece são processos físico-químicos comuns, como sequências de nucleotídeos interagindo com enzimas, com moléculas de RNA, com outras proteínas, através de forças químicas (ligações de hidrogênio, forças de van der Waals, etc.). Chamar isso de “informação” é uma metáfora conveniente, sem dúvida, mas ela não deve nos levar a pensar que a informação seja um agente causal distinto dos processos materiais que a implementam.

Os defensores da causalidade informacional, no entanto, têm uma resposta sofisticada. Eles argumentam que o nível informacional é um nível legítimo de explicação causal, e que ele não é redutível à química de nível inferior. Mesmo que cada evento que descrevemos como “informacional” seja, em última instância, implementado por processos físicos, o que explica as regularidades biológicas não são os detalhes físicos particulares, mas sim a estrutura informacional que eles realizam. Um mesmo gene (no sentido informacional) pode ser implementado por múltiplas sequências de DNA diferentes, graças à degeneração do código genético. Da mesma forma, múltiplas configurações químicas distintas podem, em princípio, realizar a mesma informação. O que importa causalmente, para o biólogo, é a informação (a instrução para produzir determinada proteína), e não o substrato específico em que ela está inscrita. Troque o substrato (de

DNA para RNA, por exemplo), mas preserve a informação, e o efeito biológico será o mesmo. Isso sugere que a informação tem, sim, um poder causal genuíno, e que tentar reduzir toda a biologia à química seria perder esse aspecto informacional da causalidade, que talvez seja o que há de mais explicativo das teorias.

O argumento da múltipla realizabilidade na biologia conecta-se com o debate mais amplo sobre emergência. Se a causalidade informacional é irredutível à causalidade química, então a informação é um nível emergente de organização com poderes causais próprios. A emergência forte (onde os poderes causais do nível superior não derivam do nível inferior) é controversa, mas sua versão fraca, onde a descrição informacional é autônoma sem violar leis físicas, é amplamente aceita na biologia teórica.

Este debate, como se vê, conecta-se diretamente com questões filosóficas mais amplas sobre níveis de explicação e redução na ciência. Se as explicações informacionais forem autônomas e irredutíveis, então a informação possui um status causal genuíno, e os sistemas de informação que construímos não são meros repositórios passivos, mas sim agentes que efetivamente modificam o mundo. Se, por outro lado, as explicações informacionais forem apenas redescrições convenientes de processos puramente físicos, então a informação, em si mesma, não causa nada; apenas as coisas físicas causam. Essa discussão tem implicações diretas para a prática profissional. Os profissionais da informação falam o tempo todo como se a informação fosse causalmente eficaz. Dizem que "a informação influencia decisões", que "a falta de informação causa erros", que "a desinformação produz polarização social". Se a informação não pode causar nada, esses modos de falar são, no máximo, metáforas úteis, mas, em última análise, falsos. Se, contrariamente, a informação pode causar algo, então os sistemas de informação não são apenas instrumentos passivos que organizam descrições da realidade; eles são parte ativa na criação e modificação da própria realidade, através dos efeitos causais que a informação desencadeia.

Esta discussão metafísica, longe de ser um desvio, conecta-se diretamente com tudo o que vimos nos capítulos anteriores. Na antropologia, argumentamos que a cultura é real, múltipla, emergente e histórica. Na linguística, vimos que a língua é simultaneamente cognitiva, social e formal. Na matemática, nos debatemos com o estranho reino dos números, situados entre o abstrato e o concre-

to. Pois bem, na ciência da informação, a mesma complexidade ontológica se manifesta com toda a força. A informação não é puramente física (ela pode ser realizada em múltiplos substratos), nem puramente mental (ela existe inscrita em documentos independentemente de ser lida), nem puramente abstrata (ela precisa de um suporte material para existir e circular). Ela é, exatamente como a cultura, a linguagem e os números, um fenômeno multifacetado que resiste a reduções simplistas e exige de nós, seus estudiosos, uma boa dose de pluralismo conceitual.

3. Epistemologia da Ciência da Informação

Se a metafísica pergunta o que é a informação, a epistemologia pergunta como nós, seres humanos, nos relacionamos com ela do ponto de vista do conhecimento. Aplicada à ciência da informação, a epistemologia levanta algumas questões fundamentais. Qual é a relação entre informação e conhecimento? Como podemos conhecer o mundo através da informação mediada por documentos, livros e sistemas digitais? Que tipo de autoridade epistêmica esses sistemas possuem? E, de forma ainda mais básica, será que informação e conhecimento são a mesma coisa?

Essa última pergunta é o ponto de partida. A relação entre informação e conhecimento é central para a área, mas também profundamente controversa. Será que informação é conhecimento? Será que o conhecimento é apenas um tipo especial de informação, mais refinada ou mais bem processada? Ou serão ambos fenômenos de naturezas fundamentalmente distintas? Um dos modelos mais influentes, e também dos mais criticados, é a chamada hierarquia DIKW, um acrônimo para Dados, Informação, Conhecimento e Sabedoria (do inglês *Data, Information, Knowledge, Wisdom*). A ideia básica é que esses quatro conceitos formam uma pirâmide, uma progressão conceitual. Partimos dos dados brutos, que, ao serem contextualizados, se tornam informação; a informação, ao ser internalizada e compreendida, se transforma em conhecimento; e o conhecimento, quando aplicado com julgamento e experiência, atinge o nível da sabedoria. As origens precisas dessa hierarquia são um tanto obscuras, mas ela aparece em livros e artigos desde os anos 1980 e se tornou um *framework* ex-

tremamente popular, especialmente em áreas como a gestão do conhecimento e os sistemas de informação.

Os dados seriam símbolos, caracteres isolados, fatos brutos desprovidos de qualquer contexto. O número “25” é um dado. A palavra “João” é um dado. Sozinhos, eles não significam muita coisa. A informação surge quando damos contexto a esses dados, quando os organizamos e interpretamos. “João tem 25 anos” já é informação. É um dado contextualizado, que passa a ter significado. O conhecimento seria a informação que foi internalizada, compreendida, e integrada à estrutura cognitiva de uma pessoa. Não é mais apenas uma frase que se pode ler; é algo que a pessoa entende e pode relacionar com outras coisas que sabe. A sabedoria, por fim, é o conhecimento aplicado com julgamento, experiência e discernimento, para tomar decisões acertadas em situações complexas. É saber não apenas o que fazer, mas quando fazer, por que fazer, e se é a coisa certa a se fazer.

É fácil ver por que essa hierarquia é tão atraente. Ela é intuitiva, didática, e organiza conceitos de forma elegante. Ela sugere um papel claro para os profissionais da informação, a saber, o de transformar dados brutos em informação organizada, que por sua vez pode ser usada por outros (pesquisadores, cidadãos, gestores) para gerar conhecimento. O problema é que, quando examinada com lentes filosóficas mais rigorosas, a hierarquia DIKW começa a se desfazer.

As críticas são sérias e numerosas. A pesquisadora Jennifer Rowley, em um artigo influente, documentou que existem múltiplas versões da hierarquia, inconsistentes entre si, e que falta a ela um rigor conceitual básico. As transições entre os níveis são vagas. O que exatamente transforma dado em informação? A resposta comum é o “contexto”. Mas o contexto, ele mesmo, é um tipo de informação. Então estaríamos usando informação para definir informação, num raciocínio circular. O que transforma informação em conhecimento? A resposta comum é “internalização” ou “compreensão”. Mas a informação também pode ser internalizada. A distinção permanece nebulosa.

O filósofo Martin Frické (2009) foi ainda mais contundente, argumentando que a hierarquia DIKW é filosoficamente incoerente. A distinção entre dado e informação, segundo ele, é insustentável, porque a mesma entidade pode ser classificada como dado ou como informação a depender da perspectiva de

quem a observa. Um número solto pode ser um dado para um estatístico, mas informação valiosa para alguém que o reconhece como sua senha bancária. Além disso, a hierarquia comete um erro categorial grave ao tratar informação como uma "etapa" para o conhecimento. O conhecimento, na tradição filosófica, é um conceito muito mais exigente. Ele requer verdade (não se pode conhecer algo falso), crença (é um estado mental de um sujeito) e justificação (é preciso ter boas razões para crer). A informação, por sua vez, não requer nada disso. Um documento pode conter informação falsa, pode existir sem que ninguém acredite nela, e não precisa de justificação. A hierarquia, ao tratar conhecimento como uma mera extensão da informação, estaria ignorando essas diferenças cruciais. Por fim, Frické (2009) aponta que o nível da sabedoria é um conceito vago demais, sobre o qual ninguém oferece uma análise clara de o que o distinguiria do conhecimento.

O erro categorial identificado por Frické (2009) remete à distinção aristotélica entre *episteme* (conhecimento científico demonstrativo), *techne* (conhecimento técnico-produtivo) e *phronesis* (sabedoria prática). A hierarquia DIKW colapsa distinções que a filosofia clássica havia separado cuidadosamente, tratando sabedoria como mero grau superior de conhecimento proposicional, quando na verdade a *phronesis* aristotélica é uma capacidade de julgamento situado irredutível ao acúmulo de informações.

Apesar dessas críticas devastadoras, a hierarquia DIKW persiste na literatura e no ensino. Isso sugere que, mesmo sendo filosoficamente frágil, ela captura algo intuitivamente correto sobre os processos de transformação informacional. Talvez a maneira mais caridosa de interpretá-la não seja como uma hierarquia ontológica (sobre os tipos de coisas que existem) ou epistêmica (sobre os tipos de conhecimento), mas sim como uma descrição pragmática de como a informação é transformada em contextos organizacionais. Ela descreve um processo, não uma estrutura do mundo.

Diante das fraquezas da DIKW, filósofos mais rigorosos buscaram desenvolver teorias alternativas sobre a relação entre informação e conhecimento. Uma das mais influentes é a de Fred Dretske (1981), apresentada em seu livro *Knowledge and the Flow of Information*. Dretske (1981) propõe analisar o conceito de conhecimento em termos puramente informacionais, sem recorrer a noções epistêmicas não analisadas, como "justificação". A ideia central é a de

que um sinal (uma percepção, um testemunho, um dado) carrega a informação de que algo é o caso se, e somente se, a probabilidade condicional de que esse algo seja o caso, dado o sinal, for igual a 1. Em outras palavras, o sinal indica com certeza que o mundo é de determinada maneira. Se a fumaça carrega a informação de que há fogo, isso significa que, dada a presença da fumaça, a probabilidade de haver fogo é 100%. Não há margem para dúvida.

Com base nessa definição de informação, Dretske (1981) então analisa o conhecimento. Um sujeito S conhece que p se, e somente se, a crença de S de que p é causada (ou sustentada) pela informação de que p, e essa informação não é degradada no processo de transmissão até chegar a S. Em termos mais simples, conhecer é ter uma crença que se originou de uma fonte perfeitamente confiável e que chegou até nós sem sofrer distorções no caminho.

Essa teoria tem virtudes notáveis. Ela conecta a epistemologia com a teoria matemática da informação e com as ciências cognitivas de uma maneira inovadora. Ela oferece uma explicação naturalista de como o conhecimento perceptual é possível, ou seja, meus olhos captam a informação sobre o ambiente, essa informação causa em mim a crença correspondente, e eu passo a conhecer o que está diante de mim. Ela também evita, elegantemente, alguns dos problemas mais espinhosos da análise tradicional do conhecimento como "crença verdadeira justificada" (os famosos contraexemplos de Gettier (1963)).

No entanto, a teoria de Dretske (1981) enfrenta objeções sérias, que seus críticos não tardaram a apontar. A mais óbvia é que a exigência de que a informação tenha probabilidade 1 é restritiva demais. No mundo real, raramente temos certezas absolutas. A fumaça quase sempre indica fogo, mas pode ser fumaça de um gerador de efeitos especiais. Meus olhos geralmente me informam corretamente, mas podem me enganar em condições de pouca luz. Se levarmos a exigência de Dretske (1981) ao pé da letra, muito pouco do que chamamos de conhecimento de fato se qualificaria. O próprio Dretske (1981) reconheceu o problema e, em trabalhos posteriores, relaxou a condição, permitindo que a informação pudesse ser apenas altamente provável. Mas essa modificação cria um novo problema. Se a informação é apenas provável, como distinguir o conhecimento genuíno da mera crença verdadeira que pode ter sido adquirida por sorte? A distinção se torna turva. Além disso, a teoria tem dificuldade em explicar o conhecimento *a priori*, i.e., o conhecimento matemático e

lógico, que não parece ser baseado em informação causal vinda do mundo exterior. Como sabemos que 2+2=4? Não há um sinal físico que carregue essa informação até nós. A teoria de Dretske (1981), focada em causalidade e informação empírica, parece não ter como acomodar esses casos.

Apesar dessas dificuldades, a tentativa de Dretske (1981) é valiosa porque nos força a levar a sério a ideia de que o conhecimento tem bases informacionais. Mas talvez o conhecimento não seja apenas uma questão de informação individual. Talvez ele seja, em grande medida, um fenômeno social. É aqui que entra a epistemologia social, um campo que tem conexões profundas com a ciência da informação.

A ideia central da epistemologia social é que o conhecimento não é produzido, validado e transmitido apenas por indivíduos isolados, mas sim por comunidades, instituições e práticas coletivas. O conhecimento científico, por exemplo, não é algo que um único gênio descobre sozinho; é o produto de uma comunidade complexa, com seus laboratórios, periódicos, conferências, sistemas de revisão por pares e disputas de prioridade. Essa dimensão social do conhecimento é o objeto de estudo da epistemologia social.

Um dos pioneiros dessa abordagem, e figura fundamental para a ciência da informação, foi o bibliotecário e teórico Jesse Shera (1970). Já em 1951, Shera (1970) cunhou o termo "epistemologia social" e passou as décadas seguintes desenvolvendo-o como o fundamento teórico para a biblioteconomia. Em seu livro *Sociological Foundations of Librarianship* (1970), Shera (1970) argumenta que bibliotecas, arquivos e outras instituições informacionais desempenham um papel absolutamente central nos processos sociais de produção e circulação do conhecimento. Para Shera (1970), a epistemologia social vai muito além da epistemologia tradicional. Enquanto esta se pergunta como um indivíduo adquire e justifica suas crenças, aquela se pergunta como o conhecimento é produzido, distribuído, integrado e consumido através de todo o tecido social. É o estudo dos sistemas de conhecimento, não apenas das mentes individuais.

Nessa visão, as bibliotecas não são meros depósitos de livros. Elas são instituições epistêmicas fundamentais. Elas não apenas armazenam o conhecimento registrado, mas também o organizam (através de sistemas de classificação), o tornam acessível (através de catálogos e serviços de referência) e mediam a relação entre os produtores e os consumidores de conhecimento. Os bibli-

otecários, nesse quadro, são verdadeiros curadores epistêmicos. As escolhas que fazem sobre o que adquirir, como classificar, o que preservar e como disponibilizar moldam, de forma concreta, o que uma comunidade pode conhecer e como pode conhecê-lo.

A caracterização dos bibliotecários como "curadores epistêmicos" pode ser aprofundada com o conceito de "guardiões do conhecimento" (*knowledge gatekeepers*) desenvolvido na sociologia da ciência. A função de filtragem, organização e legitimação exercida pelos bibliotecários é estruturalmente análoga à função dos editores científicos e dos revisores por pares, todos operam como mediadores que regulam o que conta como conhecimento disponível para uma comunidade.

Na mesma linha, o filósofo Alvin Goldman (1999), em seu livro *Knowledge in a Social World*, desenvolveu uma epistemologia social sistemática, focada em como práticas e instituições sociais afetam a aquisição de conhecimento verdadeiro. Goldman (1999) adota uma abordagem confiabilista (*reliabilism*) de orientação veritista (*veritism*), cuja ideia central é que as práticas e instituições epistêmicas, como os sistemas de informação, os métodos científicos e o jornalismo, devem ser avaliadas de acordo com sua propensão a produzir crenças verdadeiras e a reduzir o erro. Quanto mais uma prática confiavelmente nos leva à verdade, maior é o seu valor epistêmico. É uma perspectiva que coloca a ênfase nos resultados dos processos de conhecimento, e não apenas em sua correção formal.

Um dos conceitos centrais de Goldman (1999), e que tem uma relevância imensa para a ciência da informação, é o de verificabilidade (*verifiability*). Para entender por que esse conceito é tão importante, precisamos lembrar de algo que discutimos há pouco, a saber, em sociedades complexas como as nossas, dependemos massivamente do testemunho de outros e da autoridade de instituições para a maior parte do que chamamos de conhecimento. Nenhum de nós pode verificar pessoalmente a eficácia de uma vacina, a data exata de um evento histórico ocorrido há séculos, ou a veracidade de uma notícia que circula em outro continente. Simplesmente confiamos em especialistas, em jornais, em livros, em bancos de dados. Essa confiança é inevitável e, em grande medida, perfeitamente racional. Não podemos viver em estado de verificação permanente.

No entanto, para que essa confiança não seja cega, para que ela seja uma confiança justificada e não um mero ato de fé, é essencial que os sistemas sociais de informação possuam uma propriedade que Goldman (1999) chama de verificabilidade. A verificabilidade é a propriedade que um sistema tem de permitir que, em princípio, as informações que ele veicula possam ser verificadas. Não se trata, repito, de que cada cidadão saia verificando tudo o tempo todo. Isso seria humanamente impossível. Trata-se, isso sim, de que existam mecanismos institucionais que tornem a verificação possível para aqueles que, por alguma razão, desejam ou precisam realizá-la.

Pense no que isso significa na prática. Uma biblioteca que preserva as fontes originais e documenta a proveniência de seus acervos está aumentando a verificabilidade do conhecimento ali contido. Um artigo científico que lista suas referências bibliográficas de forma clara e precisa está aumentando a verificabilidade de suas afirmações. Um arquivo público que organiza seus documentos e os torna acessíveis à consulta está aumentando a verificabilidade das decisões governamentais. Um jornal que publica suas fontes e disponibiliza os documentos em que baseou uma reportagem está aumentando a verificabilidade de sua investigação.

Por outro lado, sistemas que obscurecem as fontes, que não permitem rastrear a informação até sua origem, ou que ativamente dificultam a verificação, estão, segundo esse critério, diminuindo as capacidades epistêmicas da sociedade. Algoritmos de recomendação opacos, cujo funcionamento interno é um segredo comercial; notícias que circulam sem qualquer citação ou referência verificável; documentos oficiais que são destruídos ou cujo acesso é arbitrariamente restrito, todos esses são exemplos de práticas que reduzem a verificabilidade. Elas nos forçam a confiar sem a possibilidade de, em última instância, checar. E confiança sem possibilidade de verificação, por mais confortável que seja, é um terreno frágil para construir o conhecimento coletivo de uma sociedade que se quer democrática e livre.

O argumento de Goldman (1999) sobre verificabilidade converge com a análise habermasiana das condições de validade do discurso, segundo a qual uma afirmação é válida quando pode ser resgatada (*eingelöst*) por razões que qualquer participante competente pode examinar. A verificabilidade institucional

é a condição prática de possibilidade desse resgate discursivo, sem ela, a comunicação degenera em retórica ou propaganda.

Bibliotecas e arquivos são instituições que aumentam a verifiliabilidade da sociedade. Ao preservarem fontes originais, ao documentarem a proveniência dos documentos, ao organizarem a informação de forma que permita a verificação cruzada, elas criam as condições para que o conhecimento possa ser testado e confirmado. Um pesquisador pode verificar a afirmação de um artigo científico consultando as referências originais que estão preservadas em uma biblioteca. Um cidadão pode verificar uma decisão governamental consultando os documentos oficiais em um arquivo público. Sistemas que obscurecem as fontes, que não permitem rastrear a informação até sua origem, ou que dificultam a verificação, estão, segundo esse critério, diminuindo as capacidades epistêmicas da sociedade. Algoritmos de recomendação opacos, notícias sem citações verificáveis, documentos sem metadados adequados, todos são exemplos de práticas que reduzem a verificabilidade (ou verifiliabilidade, conforme outras traduções).

Goldman (1999) também discute a divisão do trabalho epistêmico, um fenômeno central em sociedades complexas. Diferentes pessoas e instituições se especializam em diferentes domínios do conhecimento. Não se espera que um físico quântico saiba tudo sobre a história do Brasil, nem que um historiador domine a mecânica quântica. As bibliotecas universitárias especializadas, os arquivos temáticos e as bases de dados disciplinares são a materialização institucional dessa divisão de trabalho. Os profissionais da informação, nesse contexto, ocupam uma posição peculiar, dado que são especialistas de segundo nível. Eles não são especialistas em todos os domínios representados em suas coleções (isso seria humanamente impossível), mas são especialistas em como organizar, preservar e facilitar o acesso ao conhecimento especializado produzido por outros. São os especialistas da mediação.

A epistemologia social nos força a pensar sobre o conhecimento distribuído. Muitas vezes, o conhecimento não está inteiramente na cabeça de um único indivíduo, mas sim distribuído entre múltiplos agentes. Uma equipe de pesquisa, por exemplo, pode coletivamente saber como conduzir um experimento complexo, mesmo que nenhum membro individualmente possua todas as habilidades e conhecimentos necessários. O conhecimento está nas interações, nos protoco-

los, na memória institucional da equipe. O filósofo Philip Kitcher (1990) argumentou que as comunidades científicas se organizam de forma a distribuir o trabalho cognitivo eficientemente. Cientistas se especializam, perseguem linhas de pesquisa diferentes, compartilham resultados através de publicações. Essa divisão do trabalho permite que a comunidade como um todo conheça muito mais do que qualquer indivíduo isolado poderia conhecer.

Os sistemas de informação são a infraestrutura que torna possível esse conhecimento distribuído. Repositórios institucionais permitem que pesquisadores depositem e acessem os trabalhos uns dos outros. Bases de dados bibliográficas mapeiam a literatura, criando um mapa do conhecimento distribuído. Catálogos de bibliotecas revelam onde determinado conhecimento pode ser encontrado, permitindo que pesquisadores localizem a expertise de que precisam.

Mas o conhecimento distribuído levanta questões epistemológicas espinhosas. Quando um grupo possui conhecimento distribuído, o que exatamente cada indivíduo no grupo sabe? Se o conhecimento sobre como conduzir um experimento está distribuído entre cinco pesquisadores, cada um dominando uma parte diferente, alguém ali sabe, individualmente, como conduzir o experimento? Ou o conhecimento é uma propriedade apenas do grupo como coletivo? Alguns epistemólogos sociais, como Margaret Gilbert (1987), argumentam que grupos podem ter estados epistêmicos coletivos genuínos, que não são redutíveis aos estados individuais de seus membros. Um comitê pode, por exemplo, acreditar coletivamente que uma determinada política é a melhor, mesmo que nenhum membro individualmente sustente essa crença, desde que todos concordem em adotá-la como a posição oficial do grupo. Outros filósofos, como Frederick Schmitt, são mais céticos e defendem que o conhecimento de grupo sempre se reduz ao conhecimento individual, com a adição de uma estrutura apropriada de comunicação e coordenação. O grupo sabe quando seus membros individuais possuem as partes relevantes do conhecimento e conseguem integrá-las quando necessário. Para a ciência da informação, a pergunta crucial é se os sistemas de informação são meros facilitadores do conhecimento individual distribuído, ou se podem eles mesmos ser, de alguma forma, portadores de um conhecimento genuinamente coletivo.

Há ainda uma dimensão fundamental da epistemologia que se conecta diretamente com a prática dos profissionais da informação, que é a questão do

testemunho. Grande parte do que consideramos conhecimento não é adquirido por observação direta ou raciocínio próprio, mas sim por testemunho; nós aceitamos o que outros nos dizem. Confiamos em livros, em professores, em jornais, em sites. Essa dependência do testemunho é epistêmica e moralmente problemática. Como sabemos que o testemunho é confiável? Quando estamos justificados em confiar no que alguém nos diz?

Duas grandes tradições filosóficas disputam a resposta. O reducionismo sobre o testemunho sustenta que ele não é uma fonte básica de justificação. A justificação que obtemos pelo testemunho depende sempre de outras fontes, como a percepção, a memória ou a indução. Antes de aceitar o testemunho de alguém, precisamos ter razões independentes para acreditar que essa pessoa é confiável. Essas razões podem vir da nossa experiência passada com aquela pessoa, do conhecimento que temos sobre sua reputação, ou de inferências indutivas baseadas em casos anteriores. David Hume (1748/1999), o grande filósofo escocês do século XVIII, defendeu uma versão dessa visão. Em seu famoso argumento contra os milagres, ele afirmou que só podemos confiar em um testemunho sobre um milagre se a probabilidade de que o testemunho seja falso for menor do que a probabilidade de que o milagre tenha de fato ocorrido. E isso, argumentou Hume (1748/1999), quase nunca é o caso. Percebam que para Hume (1748/1999) há uma implicação de que a confiança no testemunho precisa ser respaldada por uma avaliação probabilística que depende de evidências independentes sobre a confiabilidade das testemunhas.

O anti-reducionismo, por outro lado, defende que o testemunho é uma fonte básica de justificação, tão fundamental quanto a percepção ou a memória. Não precisamos, a cada nova informação recebida, buscar evidências independentes da confiabilidade da fonte. Estamos, por padrão, justificados em aceitar o que nos é apresentado como verdadeiro, a menos que surjam razões específicas para duvidar. O filósofo Tyler Burge (1993) defende essa posição, argumentando que as pessoas estão intituladas a aceitar como verdadeiro o que lhes é apresentado como verdadeiro. Esse *entitlement* é um direito epistêmico padrão, que pode ser derrotado se surgirem evidências contrárias (se descobrirmos que a pessoa é conhecida por mentir, ou que o que ela diz é implausível), mas que não precisa ser positivamente justificada a cada vez.

Para os profissionais da informação, esse debate tem implicações práticas diretas. Se o reducionismo estiver correto, então simplesmente fornecer informação não é suficiente. É preciso também fornecer metainformação sobre a confiabilidade das fontes. Os usuários precisam de evidências de que os documentos são autênticos, de que os autores são especialistas reconhecidos, de que as editoras são respeitáveis. Se o anti-reducionismo estiver correto, por outro lado, a informação proveniente de fontes aparentemente legítimas pode ser aceita por padrão, e a verificação só se torna necessária quando há uma razão específica para suspeitar. A prática profissional, me parece, acaba assumindo uma posição intermediária. De um lado, investimos em mecanismos que funcionam como evidência de confiabilidade, como certificações de autenticidade, sistemas de revisão por pares, selos editoriais de qualidade. Isso está alinhado com a ênfase reducionista na necessidade de evidência. De outro lado, operamos com uma presunção de que os documentos presentes em coleções respeitáveis (como as de uma biblioteca universitária) são confiáveis na ausência de razões específicas para duvidar. Isso se alinha com a presunção anti-reducionista.

Uma questão correlata é a da autoridade epistêmica. Quando é que alguém ou algo conta como autoridade sobre um determinado tópico? Em ciência, essa questão é central. Os indivíduos, muitas vezes, precisam aceitar resultados que não podem verificar pessoalmente, confiando na autoridade da comunidade científica. O filósofo John Hardwig (1985) argumentou que essa dependência epistêmica é inevitável na ciência moderna, e não é, por si só, irracional. Os sistemas de informação são os grandes mediadores dessa autoridade. Os rankings de busca do Google sinalizam autoridade (as páginas que aparecem no topo são implicitamente tratadas como mais autorizadas). Os índices de citação quantificam a autoridade acadêmica (um artigo muito citado é visto como mais influente). Selos de qualidade, certificações, endossos institucionais funcionam como marcadores de autoridade. Mas essa mediação não é neutra. Quem decide o que conta como autoridade? Os algoritmos de ranking incorporam os valores e os vieses de seus designers. As métricas de citação favorecem certos tipos de pesquisa (as chamadas “ciências duras”) em detrimento de outras. Os sistemas de informação, longe de serem canais neutros, estão ativamente en-

volvidos na construção e na concessão de autoridade epistêmica, com implicações profundas para o que as comunidades podem conhecer e valorizar.

A mediação da autoridade epistêmica por sistemas algorítmicos introduz uma assimetria estrutural nova. Enquanto as autoridades tradicionais (editores, pareceristas, bibliotecários) eram identificáveis e potencialmente responsabilizáveis, os algoritmos de ranking operam de forma opaca, distribuindo autoridade sem prestação de contas. Isso constitui o que Frank Pasquale (2015) denomina "caixa preta" epistêmica, um mecanismo de poder invisível que molda o que sabemos sem poder ser interrogado.

Isso nos leva a um ponto crucial. Os sistemas de organização do conhecimento, como as classificações bibliográficas, as taxonomias e as ontologias conceituais, são, eles mesmos, epistemologias aplicadas no sentido de que não são ferramentas neutras; eles incorporam teorias sobre a estrutura do conhecimento humano e decidem, na prática, como a informação deve ser categorizada, relacionada e, em última instância, pensada. Quando Melvil Dewey desenvolveu a Classificação Decimal de Dewey (CDD) em 1876, ele estava fazendo muito mais do que criar uma ferramenta prática para bibliotecários. Ele estava implementando uma teoria sobre a estrutura do conhecimento humano. A decisão de dividir todo o saber em dez grandes classes, e de subdividir cada classe em dez divisões, e assim por diante, reflete pressupostos epistemológicos e metafísicos profundos. O sistema de Dewey assume que o conhecimento é hierárquico, que existem disciplinas fundamentais que podem ser organizadas em uma árvore, que as categorias são mutuamente exclusivas e que essa estrutura é universal, válida para todas as culturas e todos os tempos.

O sistema de Dewey é também o produto de uma teologia específica. Dewey estudou no Amherst College em ambiente fortemente cristão protestante, e a estrutura da CDD reflete essa visão de mundo, a Teologia ocupa a primeira classe (200), e a disposição das disciplinas reflete a hierarquia do saber europeu do século XIX. A crítica ao imperialismo epistêmico da CDD não é, portanto, uma acusação retroativa anacrônica; é a detecção de pressupostos que estavam presentes na sua própria origem.

Cada uma dessas suposições é filosoficamente questionável. O conhecimento pode não ser naturalmente hierárquico, mas sim organizado em redes complexas, com conexões que atravessam fronteiras disciplinares. A própria

ideia de “disciplinas” pode ser um artefato histórico das universidades europeias do século XIX, e não uma divisão natural do saber. Uma obra pode, e muitas vezes o faz, atravessar múltiplas categorias, o que força o classificador a escolher uma delas, ocultando as outras. E diferentes culturas, com diferentes tradições intelectuais, podem organizar o conhecimento de maneiras radicalmente diferentes. Impor a estrutura de Dewey a essas culturas é, em certo sentido, uma forma de imperialismo epistêmico.

Paul Otlet (1934) e Henri La Fontaine, os criadores da Classificação Decimal Universal (CDU) em 1895, já percebiam algumas dessas limitações. Eles tentaram aprimorar o sistema de Dewey permitindo o uso de facetas, i.e., múltiplas dimensões de classificação que podem ser combinadas para representar a complexidade de um assunto. Um livro sobre “o impacto econômico da construção de ferrovias na Inglaterra do século XIX” não precisaria ser classificado em uma única categoria; ele poderia ser representado pela combinação de facetas como “economia”, “transportes”, “Inglaterra” e “século XIX”. Isso já é um reconhecimento de que o conhecimento é multidimensional, e não simplesmente hierárquico.

O bibliotecário indiano S. R. Ranganathan (1931) levou essa ideia ao extremo com sua Classificação dos Dois Pontos (*Colon Classification*). Ranganathan (1931) argumentava que assuntos complexos devem ser analisados em facetas fundamentais, que ele resumiu na fórmula PMEST (Personalidade, Matéria, Energia, Espaço e Tempo). Um livro sobre “o tratamento cirúrgico de tumores cardíacos em crianças na Índia contemporânea” poderia ser classificado combinando essas facetas de forma flexível. Essa abordagem reflete uma epistemologia muito mais sofisticada, que reconhece a multidimensionalidade e a contextualidade do conhecimento.

As famosas cinco leis da biblioteconomia de Ranganathan (1931) , formuladas em 1931, são um verdadeiro tratado filosófico sobre a natureza e o propósito das bibliotecas, embora escritas na forma de aforismos simples. A primeira lei, “livros são para usar”, é uma rejeição radical da concepção de biblioteca como um museu, onde os livros são preservados como relíquias, mas dificilmente acessados. Ranganathan (1931) afirma que o valor epistêmico dos livros só se realiza no seu uso. O conhecimento não é algo estático que se preserva; é algo dinâmico que se atualiza no ato da leitura, da consulta, da pesquisa. A

segunda e a terceira leis, “a cada leitor seu livro” e “a cada livro seu leitor”, reconhecem que as conexões entre pessoas e recursos informacionais são profundamente individualizadas. Não existe o “melhor livro” em abstrato; existem livros que são apropriados para leitores específicos, com necessidades específicas, em contextos específicos. Isso é uma assunção forte de contextualidade epistêmica, dado que o que é conhecimento relevante varia com a situação do sujeito que conhece. A quarta lei, “economize o tempo do leitor”, é um princípio pragmático que reconhece que a busca por informação tem custos (tempo, esforço, atenção). Os sistemas de informação devem ser projetados para minimizar esses custos, para que o leitor possa encontrar o que precisa com a máxima eficiência. Isso reflete uma epistemologia que valoriza não apenas a verdade, mas também a eficiência no acesso à verdade. A quinta lei, “a biblioteca é um organismo em crescimento”, é talvez a mais profunda de todas. Ela reconhece que o conhecimento não é estático, mas sim evolutivo. Novos campos do saber emergem, as fronteiras entre as disciplinas se deslocam, as necessidades dos usuários se transformam. A biblioteca, como instituição que serve ao conhecimento, precisa ser capaz de crescer e se adaptar a essas mudanças. Isso é uma rejeição de qualquer epistemologia estática e uma defesa de uma visão evolucionista e dinâmica do conhecimento.

Uma abordagem contemporânea que leva essas ideias adiante é a análise de domínio, proposta por Birger Hjørland e Hanne Albrechtsen (1995). Eles argumentam que a organização do conhecimento não deve buscar uma estrutura única e universal, mas sim ser baseada na análise dos domínios epistêmicos, isto é, das comunidades de prática que compartilham objetivos, métodos, teorias e uma linguagem comum. Biólogos organizam seu conhecimento de uma forma (taxonomia evolutiva), médicos de outra (por sistemas do corpo ou por patologias), farmacêuticos de outra (por mecanismos de ação de fármacos). Cada uma dessas organizações é apropriada para os propósitos específicos daquele domínio. Isso implica que os sistemas de classificação universais, como a CDD ou a CDU, são inerentemente problemáticos. Eles impõem uma estrutura única sobre um conhecimento que é, na realidade, estruturado de maneiras múltiplas e diversas por diferentes comunidades.

Essa abordagem pluralista, no entanto, tem seus próprios custos. Se cada domínio tem sua própria estrutura, como podemos integrar o conhecimento

através dos diferentes domínios? Como um pesquisador pode navegar entre a biologia, a química e a medicina se cada área usa um sistema de classificação diferente? Talvez a solução não seja escolher entre o universalismo e o particularismo desses domínios, mas sim adotar uma estratégia de múltiplos níveis, como, por exemplo, sistemas universais para uma integração de alto nível, e sistemas específicos de domínio para uma organização mais fina e relevante para cada comunidade. O desafio, como sempre, é encontrar o equilíbrio.

Esta longa discussão epistemológica, como se pode perceber, conecta-se diretamente com os capítulos anteriores. Na antropologia, vimos o debate entre o relativismo e o universalismo no que diz respeito ao conhecimento intercultural. Na linguística, nos confrontamos com a diversidade de métodos e tipos de evidência. Na ciência da informação, esses debates retornam, agora na forma de questões sobre a possibilidade de sistemas de classificação universais e sobre a legitimidade de diferentes abordagens metodológicas para organizar o conhecimento. E a resposta que emerge, mais uma vez, é a do pluralismo metodológico. Diferentes métodos de organização do conhecimento são apropriados para diferentes propósitos. Não há, e provavelmente nunca haverá, uma única maneira "correta" de classificar todo o conhecimento humano. A tarefa do profissional da informação, longe de ser a de um mero aplicador de regras, é a de um tomador de decisões reflexivo, que precisa escolher, com base em seu julgamento informado, o método mais adequado para o contexto e os propósitos específicos em que está atuando.

4. Ética da Ciência da Informação

A ética, em seu sentido mais amplo, investiga o que devemos fazer, quais valores devemos buscar e como devemos tratar uns aos outros. Quando aplicada à ciência da informação, a ética nos força a fazer perguntas incômodas e fundamentais, como perguntar sobre quais responsabilidades temos ao criar, organizar, preservar e disseminar informação. Como podemos equilibrar valores que frequentemente entram em conflito, como a defesa da liberdade intelectual e a proteção da privacidade, ou a promoção do acesso universal e a necessidade de justiça distributiva?

Uma das tentativas mais ambiciosas e sistemáticas de responder a essas perguntas é a Ética da Informação desenvolvida pelo filósofo Luciano Floridi (1999). Apresentada inicialmente em seu artigo *Information Ethics* (1999) e depois desenvolvida em diversas obras, a teoria de Floridi (1999) propõe uma mudança radical de perspectiva. Enquanto as éticas tradicionais (como a utilitarista, a deontológica ou a das virtudes) estão centradas no ser humano e em seus interesses, a Ética da Informação adota uma perspectiva que Floridi (1999) chama de informacional-ambiental. Nessa visão, a entidade básica de consideração moral não é o ser senciente ou consciente, mas sim qualquer entidade informacional, qualquer estrutura organizada que possua uma integridade informacional.

Os conceitos centrais dessa teoria são fascinantes. Floridi (1999) introduz a noção de *infosfera*, o ambiente informacional total, análogo à biosfera. A infosfera não inclui apenas a informação digital que trafega na internet, mas todo e qualquer ambiente informacional, como documentos físicos, livros, organismos vivos, sistemas sociais. Tudo o que tem uma identidade informacional (uma estrutura, uma organização, padrões recognoscíveis) é uma entidade informacional. Pessoas, evidentemente, são entidades informacionais. Mas também o são documentos, organizações, ecossistemas, obras de arte.

O que importa moralmente, para Floridi (1999), não é a consciência ou a senciência (a capacidade de sentir dor ou prazer), mas sim a integridade informacional da entidade. O mal moral fundamental, nessa ética, é a entropia, entendida não no sentido termodinâmico estrito, mas como a degradação da estrutura informacional, o aumento da desordem, a corrupção dos dados, a destruição de sistemas organizados. Ações que aumentam a entropia na infosfera (destruir um documento único, corromper uma base de dados, desestruturar um ecossistema) são moralmente problemáticas. Ações que mantêm ou criam estrutura informacional (preservar um arquivo histórico, organizar uma biblioteca, facilitar a comunicação entre pessoas) são moralmente louváveis.

A partir dessa base, Floridi (1999) deriva princípios fundamentais, a saber, (1) a entropia não deve ser causada na infosfera, i.e., não destrua, não corrompa, não desestruture informação desnecessariamente; (2) a entropia deve ser prevenida na infosfera, i.e., proteja a informação contra a degradação, preserve estruturas informacionais valiosas; (3) a entropia deve ser removida da infosfe-

ra, i.e., restaure, organize, melhore a qualidade informacional sempre que possível; e (4) o florescimento informacional deve ser promovido, i.e., enriqueça a infosfera, crie novas estruturas informacionais, facilite fluxos informativos positivos. Esses princípios fornecem um quadro conceitual para avaliar uma vasta gama de práticas informacionais. A preservação de arquivos históricos, por exemplo, é uma forma de prevenir a entropia. A organização de bibliotecas remove entropia ao estruturar o que antes era caótico. A criação de repositórios de acesso aberto promove o florescimento informacional, ao enriquecer a infosfera com novos recursos e ao facilitar o fluxo de conhecimento.

A teoria de Floridi (1999), no entanto, não está imune a críticas. Alguns argumentam que, ao estender a consideração moral a todas as entidades informacionais, ela acaba por diluir o conceito de moralidade. Destruir um vírus de computador malicioso é, segundo a teoria, reduzir a entropia e, portanto, seria moralmente bom. Isso parece intuitivamente correto. O problema talvez seja outro, a saber, o de que a teoria precisa explicar por que certas estruturas informacionais importam moralmente mais do que outras. Um ser humano, com sua complexidade informacional e sua história, parece ter um valor muito maior do que um único arquivo digital. Floridi (1999) precisa de critérios para fazer essas distinções. Além disso, os quatro princípios podem facilmente entrar em conflito. Preservar toda a informação, sem critérios de seleção, pode impedir a criação de nova informação, que muitas vezes requer a superação ou a substituição da velha. A teoria não oferece um método claro para resolver esses conflitos.

A crítica à ética de Floridi (1999) pode ser aprofundada com o conceito de incomensurabilidade de valores, i.e., mesmo aceitando que tanto a preservação de documentos históricos quanto a proteção de dados pessoais são valores informacionais, não há uma escala comum que permita compará-los e decidir qual deve prevalecer num conflito. A teoria florideana, ao tratar a entropia como o mal fundamental, pressupõe implicitamente uma escala de mensurabilidade que a realidade dos conflitos de valores desmente.

Apesar dessas críticas, a Ética da Informação de Floridi (1999) oferece uma perspectiva extremamente valiosa, especialmente para os profissionais da informação. Ela ressalta a importância moral intrínseca do trabalho de preservação, organização e acesso à informação. Ela justifica, em termos éticos, o inves-

timento social em bibliotecas, arquivos e sistemas de informação, vendo-os não como meras utilidades práticas, mas como verdadeiros imperativos morais em uma era informacional.

Independentemente de aceitarmos ou não a teoria de Floridi (1999) em sua totalidade, podemos reconhecer que a informação possui um valor moral que vai além de sua utilidade instrumental. Várias razões sustentam essa afirmação. Uma delas é que a informação é necessária para a autonomia. Agentes autônomos são aqueles que fazem escolhas informadas, baseadas em uma compreensão adequada das opções disponíveis e das consequências de cada uma. Sem acesso à informação relevante e confiável, a autonomia fica severamente comprometida. Se não posso saber quais são os efeitos colaterais de um medicamento, não posso consentir de forma autônoma em tomá-lo. Se não tenho acesso a informações sobre os candidatos em uma eleição, meu voto não é uma expressão genuína de minha vontade, mas um ato de fé cego. Para filósofos como Kant (1785/2005), a autonomia é o fundamento da dignidade humana. Se a informação é condição de possibilidade da autonomia, então garantir o acesso à informação é uma exigência moral derivada do próprio respeito à dignidade humana.

Além disso, a informação é também necessária para a justiça. O filósofo John Rawls (1971), em sua obra *Uma Teoria da Justiça*, argumenta que uma sociedade justa é aquela cujas desigualdades podem ser justificadas a todos os seus membros sob um "véu de ignorância", uma situação hipotética em que ninguém conhece sua posição social futura. No mundo real, no entanto, não há véu de ignorância. Alguns têm acesso privilegiado a informações cruciais, enquanto outros são mantidos na ignorância. Essas assimetrias informacionais criam e perpetuam desigualdades de poder que são, em si mesmas, injustas. O acesso equitativo à informação, portanto, é um requisito básico de justiça social. Sociedades que permitem que elites concentrem a informação enquanto as massas permanecem desinformadas estão violando princípios fundamentais de justiça.

Grande parte da informação tem as características econômicas de um bem público. No jargão da economia, um bem é "não-rival" quando o uso que uma pessoa faz dele não diminui a quantidade disponível para outras. Minha leitura de um artigo científico não impede que você leia o mesmo artigo. Um bem é

“não-excludente” quando é difícil (ou custoso) impedir que alguém tenha acesso a ele. Uma vez que um conhecimento científico é publicado, é muito difícil impedir que ele se espalhe. O conhecimento científico, o patrimônio cultural e os registros históricos são bens públicos por excelência. Eles beneficiam a sociedade como um todo. O problema é que os mercados, deixados à sua própria dinâmica, tendem a subproduzir bens públicos, porque os produtores individuais não conseguem capturar todo o valor social que geram. Daí a necessidade de ação coletiva para garantir a provisão desses bens. Instituições como bibliotecas públicas, arquivos nacionais e políticas de acesso aberto são a materialização dessa necessidade.

A análise econômica da informação como bem público tem um complemento filosófico importante, o de que a não-rivalidade da informação implica que sua distribuição não cria escassez, ao contrário dos bens materiais, partilhar informação não a diminui. Isso fundamenta uma crítica filosófica ao regime de propriedade intelectual, pois ao tratar a informação como bem privado e excludente através de patentes e direitos autorais, o sistema cria artificialmente a escassez que a natureza da informação não exige.

Por fim, a informação conecta passado, presente e futuro. Os documentos históricos preservam a memória coletiva de uma sociedade, permitindo que as gerações futuras aprendam com os acertos e erros do passado. Os registros que produzimos hoje serão as fontes primárias dos historiadores do amanhã. A informação é o medium através do qual as civilizações acumulam conhecimento ao longo das gerações. A responsabilidade de preservar esse patrimônio informacional é, portanto, uma responsabilidade intergeracional. Nossas decisões sobre o que preservar e como preservar terão consequências profundas para aqueles que virão depois de nós.

Dentro desse vasto campo ético, um valor se destaca como central para os profissionais da informação, o da liberdade intelectual. Trata-se do direito de buscar, receber e transmitir informações e ideias livremente, sem censura ou restrições arbitrárias. Mas esse valor, tão caro, frequentemente entra em conflito com outros valores igualmente legítimos. Como navegar por essas águas turbulentas?

A defesa clássica da liberdade de expressão e de pensamento, e que continua a influenciar profundamente o debate, é a de John Stuart Mill (1859/1991),

em seu ensaio *Sobre a Liberdade* (*On Liberty*), de 1859. Mill (1859) oferece vários argumentos poderosos. Há o argumento da verdade, de que mesmo as opiniões que nos parecem falsas ou ofensivas devem ser permitidas, em primeiro lugar, porque nossa própria certeza pode estar errada (somos falíveis), e a história está cheia de exemplos de opiniões consideradas heréticas que se revelaram verdadeiras (pense em Galileu). Em segundo lugar, mesmo uma opinião totalmente falsa pode conter um grão de verdade que, de outra forma, seria ignorado. Em terceiro lugar, o confronto constante com o erro é o que mantém viva e vigorosa a compreensão da verdade. Sem oposição, a verdade se torna um dogma morto, repetido mecanicamente, mas não verdadeiramente compreendido. Há também o argumento do debate, de que as verdades que nunca são contestadas perdem sua vitalidade. Para que permaneçam significativas, precisam ser constantemente defendidas contra objeções. A censura não apenas suprime o erro, mas também empobrece a verdade. E, finalmente, o argumento da autonomia, de que os indivíduos têm o direito fundamental de formar suas próprias opiniões e fazer suas próprias escolhas, mesmo que essas escolhas estejam erradas. O paternalismo que protege as pessoas de ideias perigosas as trata como crianças, não como agentes autônomos.

Mill (1859/1991), no entanto, não era um defensor da liberdade absoluta. Ele reconhecia um limite fundamental, expresso em seu famoso princípio do dano (*harm principle*), segundo o qual a única justificativa para restringir a liberdade de um indivíduo é evitar que ele cause dano a outros. Aplicado à liberdade de expressão, isso significa que discursos que causam dano direto e iminente podem, sim, ser legitimamente restritos.

Mas o que conta como "dano relevante"? Os filósofos distinguem várias dimensões. Há o dano direto e indireto. Incitar uma multidão à violência iminente contra uma pessoa é um dano direto. Expressar uma opinião racista que, ao longo do tempo, contribui para um clima de discriminação, causa um dano apenas indireto. Tipicamente, apenas o dano direto justifica a restrição. Há também o dano físico e o dano psicológico. O discurso de ameaça de violência física causa um dano claro. O discurso ofensivo que causa angústia psicológica também é um dano, mas é muito mais difícil de avaliar e justificar uma restrição com base nele, pois pessoas razoáveis podem discordar profundamente sobre o que é meramente ofensivo e o que é intolerável. Há, ainda, o dano a indivíduos e o

dano à sociedade. A difamação prejudica um indivíduo específico e identificável. Um discurso que, supostamente, corrompe a moral pública ou ofende valores sociais compartilhados[2] causa um dano difuso e de difícil especificação.

Essas distinções são úteis, mas a aplicação delas a casos concretos é quase sempre controversa. Consideremos o discurso de ódio, i.e., expressões que diminuem grupos sociais com base em raça, religião, etnia, orientação sexual ou gênero. Tais discursos causam dano psicológico genuíno aos membros dos grupos-alvo e podem contribuir para um clima social de discriminação e violência. Mas eles também são, inegavelmente, uma forma de expressão de opinião, por mais repugnante que seja. As democracias liberais divergem profundamente sobre como lidar com isso. Países europeus, como a Alemanha e a França, criminalizam o discurso de ódio de forma mais ampla. Os Estados Unidos, com sua tradição de forte proteção à Primeira Emenda, só permitem a restrição em casos extremos, como quando o discurso incita violência iminente e específica.

Consideremos também a desinformação (ou *fake news*). A informação falsa, disseminada intencionalmente, pode causar danos graves quando influencia decisões importantes, como no caso da saúde pública (movimentos antivacina) ou da política (interferência em processos eleitorais). No entanto, quem determina, com autoridade, o que conta como desinformação? A história está repleta de exemplos em que a ortodoxia científica ou política estava errada, e a dissidência, perseguida, estava certa. "Censurar a desinformação" pode facilmente se tornar uma ferramenta para silenciar a dissidência legítima.

Assim, para os profissionais da informação, esses dilemas não são abstratos. Eles se manifestam em decisões cotidianas. Uma biblioteca pública deve adquirir e disponibilizar um livro que contém discurso de ódio? Um arquivo histórico deve preservar panfletos de propaganda genocida do passado, mesmo que seu conteúdo seja abominável? Uma biblioteca universitária deve catalogar material pornográfico que faça parte de uma coleção especial? Não há respostas fáceis.

---

[2] Caso você queira saber mais sobre esse assunto, leia meu texto "O direito de não ser ridicularizado" publicado no livro que organizei em 2020, chamado *Refutações*, a partir de múltiplas discussões entre um agradabilíssimo grupo de filósofos discutindo temas práticos e atuais, sobre ética, política, estética e economia, por meio de ensaios curtos (https://www.editorafi.org/09refuta).

Os bibliotecários, para lidar com essas questões, desenvolveram uma distinção importante, embora nem sempre fácil de aplicar na prática, entre seleção e censura. A seleção é vista como uma escolha profissional positiva, i.e., o bibliotecário decide adquirir ou incluir um material com base em critérios legítimos, como sua qualidade literária ou acadêmica, sua relevância para a missão da instituição e para as necessidades da comunidade atendida, a necessidade de equilibrar a coleção representando múltiplos pontos de vista, a demanda dos usuários e, claro, as limitações orçamentárias. A censura, por outro lado, é vista como uma exclusão negativa, i.e., a decisão de não adquirir ou de remover um material baseada na desaprovação de seu conteúdo, por razões ideológicas, políticas, religiosas ou morais pessoais.

A distinção entre seleção e censura tem fundamento na distinção entre atos de omissão (não adquirir) e atos de comissão (remover o que já está). Filosoficamente, essa distinção não é absoluta, omissões podem ser tão moralmente relevantes quanto ações, mas ela tem um valor heurístico, pois orienta a atenção para a intencionalidade e a motivação do ato, que são os fatores moralmente decisivos.

A distinção visa a preservar o profissionalismo e, ao mesmo tempo, respeitar o princípio da liberdade intelectual. Reconhece-se que o bibliotecário não pode adquirir tudo o que é publicado; escolhas precisam ser feitas. Mas se essas escolhas forem pautadas por critérios profissionais transparentes e publicamente defensáveis, e não por preferências ideológicas pessoais ou por pressões externas, então a integridade intelectual da instituição estará preservada. Os críticos, no entanto, apontam que a distinção é mais frágil do que parece. Toda seleção, argumentam, inevitavelmente reflete valores. O que conta como “qualidade” ou “relevância” não é um dado objetivo, mas um julgamento que incorpora uma visão de mundo. Reconhecer isso, porém, não significa que a distinção entre seleção e censura deva ser abandonada; mas antes que ela precisa ser aplicada com transparência e humildade. Os bibliotecários devem ser capazes de articular claramente os critérios que orientam suas escolhas, devem estar dispostos a defendê-los publicamente e devem manter-se abertos a revisá-los quando novas informações ou argumentos assim o justificarem. Falamos aqui sobre bibliotecários, mas as conclusões são expansíveis para qualquer cientista da informação.

Outro valor fundamental, e que enfrenta desafios crescentes na era digital, é a privacidade informacional. Definida pelo pioneiro Alan Westin (1967) como o direito dos indivíduos de determinar quando, como e em que medida a informação sobre eles é comunicada a outros, a privacidade é um direito fundamental reconhecido nas democracias liberais. Mas por que ela importa moralmente? Várias justificativas foram propostas. A autonomia requer um espaço privado onde possamos ser autênticos, vulneráveis e experimentar sem a pressão constante do olhar público. A dignidade, na tradição kantiana, exige que as pessoas sejam tratadas como fins em si mesmas, e não como meros objetos. A vigilância não consentida, ao nos reduzir a pontos de dados a serem monitorados e manipulados, viola essa dignidade. As relações íntimas requerem revelação seletiva e confiança; a privacidade permite que construamos gradações de proximidade. O poder é outra dimensão crucial, pois informação é poder, e organizações ou governos que acumulam vastas quantidades de dados pessoais adquirem um poder desproporcional sobre os indivíduos. A privacidade é um antídoto contra esse desequilíbrio. Finalmente, há o argumento do direito ao esquecimento ou da segunda chance. Pessoas mudam, aprendem, evoluem e, às vezes, erram. Se todos os erros do passado forem permanentemente registrados e facilmente acessíveis, a capacidade de recomeçar e de se reinventar fica severamente comprometida. A privacidade é uma condição necessária para o crescimento pessoal.

As bibliotecas têm uma longa e orgulhosa tradição de proteger a privacidade de seus usuários. O código de ética da American Library Association (ALA, 1995), por exemplo, é explícito ao afirmar que se deve proteger o direito de cada usuário à privacidade e à confidencialidade com relação à informação buscada ou recebida, e aos recursos consultados, emprestados ou adquiridos. Os registros de empréstimo podem revelar informações altamente sensíveis sobre interesses, crenças religiosas ou políticas, opções sexuais, problemas de saúde. Um cidadão deve poder ler sobre qualquer assunto, por mais controverso que seja, sem medo de ser exposto ou retaliado.

O ambiente digital, no entanto, complicou drasticamente essa missão. Sistemas digitais geram, automaticamente, logs detalhados de cada interação, registrando quem acessou o quê, quando, de onde, por quanto tempo, em que ordem. Algoritmos de recomendação analisam o comportamento dos usuários

para personalizar serviços. Análises de big data permitem que bibliotecas entendam padrões de uso em larga escala. Essas capacidades têm valor legítimo para a gestão e o aprimoramento dos serviços. Mas elas também criam enormes riscos para a privacidade.

Diante disso, alguns princípios éticos para a gestão da privacidade em ambientes digitais têm sido propostos e amplamente debatidos. A minimização de dados recomenda coletar apenas os dados estritamente necessários para a função legítima que se pretende executar. A transparência exige que os usuários sejam informados de forma clara e acessível sobre quais dados estão sendo coletados, como serão usados, por quanto tempo serão retidos e com quem poderão ser compartilhados. O consentimento informado e genuinamente voluntário deve ser obtido antes da coleta de dados sensíveis, e os usuários não devem ser coagidos a concordar como condição para ter acesso a serviços básicos. A segurança implica a adoção de medidas técnicas e organizacionais robustas para proteger os dados contra acessos não autorizados, como criptografia, controles de acesso e auditorias regulares. A retenção limitada determina que os dados devem ser apagados assim que não forem mais necessários para a finalidade para a qual foram coletados. O propósito limitado estabelece que os dados só podem ser usados para os fins para os quais foram coletados, e não podem ser compartilhados com terceiros sem novo consentimento. Esses princípios, que estão na base de marcos regulatórios como o GDPR europeu, expressam compromissos éticos profundos que os profissionais da informação precisam internalizar e defender.

A ameaça à privacidade, no entanto, não vem apenas de sistemas mal configurados. Ela é constitutiva do próprio modelo de negócio que domina a internet. A pesquisadora Shoshana Zuboff (2019), em seu livro *A Era do Capitalismo de Vigilância* (*The Age of Surveillance Capitalism*), cunhou um termo que se tornou indispensável para entender a dinâmica contemporânea. O capitalismo de vigilância é um novo tipo de ordem econômica que trata a experiência humana como matéria-prima gratuita para a extração de dados. Plataformas digitais como Google e Facebook rastreiam cada clique, cada pesquisa, cada compra, cada localização, cada interação social. Algoritmos de aprendizado de máquina processam esses dados em escala massiva, identificando padrões e construindo perfis preditivos extremamente precisos sobre o comportamento futuro dos

usuários. Essa capacidade preditiva é então vendida como um produto a anunciantes, empregadores, seguradoras e governos. Zuboff (2019) argumenta que isso representa uma ameaça existencial à autonomia humana. Quando nosso comportamento é continuamente monitorado e moldado por algoritmos de manipulação, deixamos gradualmente de ser agentes autônomos e nos tornamos objetos de controle, peças em uma engrenagem econômica que nos ultrapassa.

Para os profissionais da informação, as implicações são profundas. Bibliotecas que oferecem acesso a plataformas de e-books, por exemplo, podem estar, sem saber, expondo os hábitos de leitura de seus usuários a um nível de detalhe que seria impensável no mundo dos livros físicos (que páginas foram lidas, quanto tempo em cada página, quais trechos foram destacados). As bibliotecas precisam resistir ativamente às demandas dos fornecedores por esses dados granulares. Análises de uso podem e devem ser feitas, mas de formas que preservem a privacidade, como através de agregação, anonimização ou uso de dados sintéticos. Os arquivistas digitais, por sua vez, precisam considerar cuidadosamente quais metadados preservar e quais remover, já que informações como data e hora de criação, modificação e acesso de um arquivo podem revelar detalhes sensíveis sobre a vida das pessoas. Acima de tudo, os profissionais da informação têm a responsabilidade de educar os usuários sobre os riscos à privacidade no ambiente digital e sobre as estratégias disponíveis para se protegerem. Promover a literacia digital é também promover a consciência sobre a vigilância.

Para além da privacidade, a questão da justiça no acesso à informação é central para a ética da área. A informação é distribuída de forma profundamente desigual no mundo e nas sociedades. Enquanto alguns têm acesso amplo e irrestrito a uma infinidade de recursos, outros sequer têm acesso às condições mais básicas. Essas desigualdades levantam questões urgentes de justiça.

O Artigo 19 da Declaração Universal dos Direitos Humanos é inequívoco, ao dizer que todo ser humano tem direito à liberdade de opinião e expressão, o que inclui a liberdade de, sem interferência, ter opiniões e de procurar, receber e transmitir informações e ideias por quaisquer meios, independentemente de fronteiras. Isso significa que o acesso à informação não é um mero privilégio, mas sim um direito humano fundamental. Na prática, no entanto, o acesso é profundamente desigual em múltiplas dimensões. Há a desigualdade econômi-

ca, dado que a informação custa dinheiro. Livros, assinaturas de periódicos, acesso a bases de dados, conectividade de internet de qualidade são bens caros, inacessíveis para grande parte da população mundial. Há também a desigualdade geográfica, pois pessoas em áreas rurais ou remotas têm muito menos acesso a bibliotecas, arquivos e infraestrutura de internet de alta velocidade. A infraestrutura informacional tende a se concentrar nos grandes centros urbanos. Há a desigualdade educacional, porque a literacia (a capacidade de ler, escrever e compreender textos) e a literacia digital (a capacidade de usar tecnologias digitais de forma eficaz e crítica) são desigualmente distribuídas. O analfabetismo funcional e o analfabetismo digital excluem grandes contingentes populacionais do acesso efetivo à informação. Há a desigualdade linguística, já que a maior parte do conteúdo científico, acadêmico e cultural disponível online está em inglês. Falantes de línguas menos dominantes têm seu acesso severamente limitado. As traduções são escassas, caras e, muitas vezes, de baixa qualidade. E há, finalmente, uma desigualdade de poder no fato de que as elites políticas e econômicas têm acesso a informações privilegiadas, não disponíveis ao público em geral. Essas assimetrias informacionais não são um mero acidente; elas criam e perpetuam as próprias desigualdades de poder.

John Rawls (1971), como vimos, argumenta que as desigualdades sociais só são justas se beneficiarem também os menos favorecidos. Aplicando esse princípio à informação, as desigualdades de acesso só seriam justificáveis se os arranjos que as produzem também beneficiassem, de alguma forma, aqueles que têm menos acesso. Essa condição, como se pode imaginar, raramente é satisfeita na realidade. As desigualdades informacionais, tipicamente, beneficiam os já privilegiados às custas dos desfavorecidos. Daí a conclusão de que a justiça exige redistribuição, i.e., políticas ativas que ampliem o acesso para os grupos marginalizados.

O fenômeno da exclusão digital (ou *digital divide*) é a manifestação mais visível dessas desigualdades no mundo contemporâneo. Mas é importante notar que a exclusão digital não é um fenômeno monolítico. Podemos distinguir pelo menos três níveis. A primeira divisão é a mais básica, a saber, o acesso ou não às tecnologias digitais fundamentais, como um computador e uma conexão à internet. A segunda divisão se refere às diferenças na qualidade do acesso (banda larga rápida versus conexão discada lenta) e nas competências de uso

(usuários que sabem usar as ferramentas de forma sofisticada versus usuários que mal sabem navegar). A terceira divisão diz respeito à capacidade de usar as tecnologias para obter ganhos tangíveis e significativos na vida real, como conseguir um emprego melhor, acessar serviços de saúde e educação de qualidade, ou participar ativamente da vida política.

A exclusão digital não é uma mera inconveniência; ela é uma injustiça por várias razões. Ela limita oportunidades, já que o acesso ao emprego, à educação superior e a serviços governamentais cada vez mais exige algum nível de inclusão digital. Ela impede a participação democrática, uma vez que a democracia contemporânea requer cidadãos informados, e a esfera pública é, em larga medida, uma esfera digital. E ela perpetua desigualdades, já que a exclusão digital está fortemente correlacionada com desigualdades preexistentes (pobreza, racismo, marginalização geográfica) e as reforça, criando um ciclo vicioso de exclusão.

A filósofa Fricker (2007) desenvolveu um conceito que ilumina de forma poderosa essas dinâmicas, o de injustiça epistêmica. Trata-se de um tipo específico de injustiça que prejudica alguém precisamente em sua capacidade como conhecedor, como produtor ou transmissor de conhecimento. Fricker (2007) distingue duas formas principais. A injustiça testemunhal ocorre quando se dá menos credibilidade ao testemunho de alguém devido a um preconceito identitário. Mulheres, pessoas negras, pobres, LGBTQIAPN+ frequentemente têm seu testemunho desacreditado ou simplesmente ignorado, não por falta de evidência, mas por causa de quem são. A injustiça hermenêutica ocorre quando alguém carece dos recursos conceituais necessários para articular e dar sentido a suas próprias experiências sociais. Antes que o conceito de “assédio sexual” fosse cunhado e se tornasse de domínio público, as mulheres que sofriam essa violência não tinham as palavras para descrever o que lhes acontecia, o que as impedia de serem compreendidas e de reivindicar justiça. A exclusão digital, nesse quadro, pode ser vista como uma forma contemporânea de injustiça epistêmica. Ao privar pessoas de acesso à informação e às ferramentas para produzi-la e compartilhá-la, a exclusão digital as constitui como “conhecedores de segunda classe”, incapazes de participar plenamente das trocas epistêmicas que definem a vida social.

Diante de tudo isso, fica claro que os profissionais da informação não são, e não podem se considerar, técnicos neutros. Eles são agentes morais com responsabilidades sociais profundas, que derivam do papel especial que ocupam nas ecologias informacionais. Eles exercem uma curadoria epistêmica, dado que, ao selecionar, organizar, preservar e facilitar o acesso ao conhecimento, suas decisões moldam, de forma concreta, o que as comunidades podem conhecer. Têm, portanto, a responsabilidade de exercer essa curadoria de forma a promover o bem epistêmico da coletividade, e não apenas interesses particulares ou de grupos restritos. Eles atuam como mediadores entre os produtores do conhecimento (pesquisadores, autores, cientistas, artistas) e os consumidores (estudantes, profissionais, cidadãos). Essa mediação deve ser exercida com imparcialidade, sem favorecer arbitrariamente certas vozes em detrimento de outras, e com o objetivo de tornar o conhecimento acessível a todos que dele necessitam. Eles são defensores do acesso. Como guardiões de recursos informacionais valiosos, têm a responsabilidade especial de defender o acesso contra restrições injustas, sejam elas a censura governamental, os preços abusivos de editoras comerciais ou a vigilância invasiva que desencoraja a consulta. Eles são educadores informacionais. Sua função não se limita a fornecer informação; eles devem também educar os usuários para que se tornem consumidores críticos e produtores competentes de informação. Ensinar a avaliar fontes, a identificar a desinformação, a buscar de forma eficiente e a usar a informação de maneira ética é uma responsabilidade social profunda, especialmente em tempos de pós-verdade e polarização. Eles são preservadores do patrimônio. Arquivistas e bibliotecários, em particular, custodiam a memória coletiva e o patrimônio cultural da humanidade. Essa função não beneficia apenas os contemporâneos; ela é uma responsabilidade para com as gerações futuras, que terão o direito de conhecer o seu passado.

A American Library Association codifica algumas dessas responsabilidades em seu código de ética, afirmando que os profissionais devem fornecer o mais alto nível de serviço a todos os usuários, por meio de recursos e serviços apropriados, precisos, imparciais e corteses; devem defender os princípios da liberdade intelectual e resistir a esforços de censura; e não devem avançar interesses privados às custas dos usuários, dos colegas ou da instituição empregadora.

Há, por fim, uma dimensão política inescapável no trabalho do profissional da informação. John Dewey (1916), o grande filósofo e educador americano, argumentou que a educação é essencial para a democracia. Cidadãos democráticos precisam ser capazes de pensar criticamente, avaliar argumentos, formar opiniões fundamentadas e participar da deliberação pública. Dewey (1916) via as bibliotecas públicas como instituições democráticas fundamentais, espaços que educam os cidadãos e fornecem os recursos para uma participação informada.

Essa visão permanece, ou talvez se torne ainda mais, relevante em nosso tempo. A democracia contemporânea enfrenta crises profundas de desinformação, polarização e desconfiança das instituições. As bibliotecas desempenham um papel crucial nesse cenário. Elas oferecem espaços neutros em sociedades profundamente divididas, onde pessoas de diferentes origens e opiniões podem interagir, encontrar múltiplas perspectivas e engajar-se em conversas civilizadas. Elas promovem a alfabetização informacional, ensinando as habilidades essenciais para navegar no turbilhão de informações, identificar fontes confiáveis e pensar criticamente sobre o que se lê e se vê. Elas fornecem acesso equitativo a recursos que, de outra forma, estariam fora do alcance de muitos, reduzindo as desigualdades informacionais que minam a democracia. Elas preservam a memória através de arquivos que tornam os governos responsáveis (*accountable*), permitindo que os cidadãos verifiquem afirmações oficiais e contestem narrativas dominantes. E funcionam como um fórum público, hospedando debates, palestras, exposições e outros eventos que facilitam a deliberação democrática sobre questões de interesse coletivo.

Essa dimensão política não significa, em absoluto, que as bibliotecas ou os arquivos devam ser partidários em disputas eleitorais. Pelo contrário, devem ser escrupulosamente imparciais, representando uma pluralidade de perspectivas. Mas a imparcialidade não é o mesmo que neutralidade sobre os valores democráticos fundamentais. As bibliotecas são, e devem ser, comprometidas com valores como o acesso equitativo, o pluralismo de ideias, a racionalidade do debate público e a liberdade intelectual. Esses são os valores que sustentam a própria possibilidade de uma democracia liberal saudável, e é em nome deles que os profissionais da informação são chamados a agir.

5. Considerações Finais

Ao longo deste capítulo, percorremos um caminho que nos levou às questões mais fundamentais sobre a ciência da informação. Na metafísica, vimos que a informação tem uma natureza dual e complexa, sendo simultaneamente física e semântica, e que diferentes teorias (realismo, construtivismo, a visão radical de Wheeler, 1990) são apropriadas para iluminar diferentes aspectos desse fenômeno multifacetado. A informação resiste a reduções simplistas e exige de nós, ao menos pragmaticamente, uma postura de pluralismo ontológico.

Na epistemologia, exploramos as relações intrincadas entre informação e conhecimento, desde a popular (mas frágil) hierarquia DIKW até as teorias mais rigorosas de Dretske (1981) e as contribuições da epistemologia social de Shera (1970) e Goldman (1999). Vimos como instituições informacionais mediam a autoridade epistêmica, como os sistemas de classificação são, eles mesmos, epistemologias aplicadas, e como a questão do testemunho e da confiança é central para a prática profissional. Mais uma vez, o pluralismo metodológico se mostrou a resposta mais adequada, dado que diferentes tipos de conhecimento e diferentes propósitos exigem diferentes métodos de organização e acesso.

Na ética, finalmente, nos confrontamos com os dilemas mais urgentes da prática profissional. Desde a Ética da Informação de Floridi (1999), passando pela defesa da liberdade intelectual em Mill (1859/1991), até as questões prementes sobre privacidade na era do capitalismo de vigilância, e a justiça no acesso à informação, vimos que o profissional da informação não pode se esconder atrás de uma pretensa neutralidade técnica. Suas decisões têm consequências profundas para a autonomia, a dignidade e a participação democrática dos cidadãos.

O pluralismo pragmático que emerge de toda essa discussão, e que ecoa os capítulos anteriores, reconhece que a informação é um fenômeno rico demais para ser capturado por uma única teoria, um único método ou um único valor. A tarefa do profissional da informação é a de exercer um julgamento reflexivo, informado por uma compreensão filosófica clara dos conceitos em jogo, dos métodos disponíveis e dos valores em conflito. É uma tarefa exigente, sem

dúvida, mas é também o que torna a profissão tão fascinante e tão central para o funcionamento de sociedades democráticas e pluralistas.

Este capítulo, portanto, prepara o terreno de forma decisiva para os próximos. Na arquivologia, veremos como essas questões sobre a natureza dual da informação, sua autenticidade e sua organicidade se manifestam de forma ainda mais específica no contexto dos documentos de arquivo. Na biblioteconomia, veremos como os livros e as coleções são organizados para o acesso e a disseminação, colocando em prática os princípios epistemológicos e éticos que discutimos. Na museologia, veremos como os objetos se tornam portadores de significado cultural, adicionando uma nova camada de complexidade à nossa compreensão da informação. Nossa trilha nesse caminho de tijolos amarelos continua, partindo da reflexão sobre a informação em si mesma em direção às instituições que a custodiam e, a partir delas, para as experiências que elas tornam possíveis. O movimento argumentativo do capítulo, da metafísica (o que é a informação) à epistemologia (como conhecemos através dela) e à ética (que responsabilidades ela gera), reproduz a estrutura tripartite da filosofia clássica e demonstra que a ciência da informação não é uma disciplina aplicada que importa conceitos filosóficos de fora, mas um campo que gera internamente problemas filosóficos de primeira grandeza. O reconhecimento dessa autonomia filosófica é uma condição extremamente necessária para o desenvolvimento de uma identidade disciplinar madura.

6

Filosofia da Arquivologia

1. Por que uma Filosofia da Arquivologia?

Quando um arquivista organiza os documentos de um fundo, quando avalia quais registros merecem ser preservados permanentemente e quais podem ser eliminados, quando define políticas de acesso a papéis sensíveis, é tentador pensar que ele está apenas executando tarefas técnicas. Seguindo um manual. Aplicando o que aprendeu nos cursos e nos treinamentos. Mas, se a gente parar para pensar com calma, cada uma dessas decisões aparentemente banais carrega questões filosóficas profundas, daquelas que a gente normalmente não vê porque está ocupado demais com o dia a dia, mas que estão ali, silenciosamente orientando tudo.

O que é, afinal, um documento de arquivo? Não é só um papel velho guardado numa caixa, nem apenas um arquivo digital perdido numa pasta. Um documento tem uma história, uma função, uma razão de ser. Ele foi criado por alguém, num determinado contexto, com uma finalidade específica. E como é que esses documentos se relacionam entre si? Eles não estão soltos, jogados aleatoriamente; eles formam conjuntos, fundos, séries. Que tipo de coisa é um fundo arquivístico? É apenas a soma dos documentos que o compõem ou existe ali uma organicidade própria, algo como uma identidade coletiva que transcende cada peça individual? A pergunta não é meramente terminológica; dela depende a justificação filosófica de princípios como o da proveniência e o da ordem original, sem os quais a arquivologia perderia sua base normativa.

E o que garante que um documento criado há cem anos, que já amarelou, que já foi restaurado, que já passou por tantas mãos, continua sendo o mesmo documento? Ele ainda é autêntico? Ainda podemos confiar nele como prova do que aconteceu no passado? E, talvez o mais espinhoso de tudo, que responsa-

bilidades temos com as pessoas que estão representadas nesses papéis? Com os criadores, com os citados, com os que tiveram suas vidas, seus negócios, suas decisões registradas ali? E com os pesquisadores do futuro, que vão depender do que a gente decidiu guardar ou jogar fora para reconstruir a história? Não há experimento científico, não há software de gestão documental, não há estatística de consulta ou índice de uso que responda a essas perguntas. Elas exigem outro tipo de trabalho. Exigem que a gente pare, desacelere, e pense sobre os conceitos que estamos usando, sobre os pressupostos que estão silenciosamente orientando cada decisão que tomamos. Exigem, em suma, filosofia.

A arquivologia, como campo profissional e acadêmico, não é nenhuma novidade. Há séculos a gente guarda documentos e reflete sobre a melhor maneira de fazer isso. No século XX, a disciplina se consolidou em torno de princípios que se tornaram praticamente sagrados para os arquivistas em todo o mundo. O princípio da proveniência (*provenance*), por exemplo, manda que os documentos vindos de uma mesma entidade não sejam misturados com os de outra. Parece óbvio, quase um mandamento de organização, mas carrega uma tese forte sobre a natureza orgânica dos conjuntos documentais. O princípio da ordem original (*original order*) diz que a organização dada aos documentos pelo seu criador deve ser respeitada. Não é um mero capricho estético ou uma mania de arrumação; é uma aposta de que aquela ordem específica carrega uma informação valiosa sobre o funcionamento da entidade que os produziu, sobre os fluxos de trabalho, as hierarquias, as prioridades. E tem a avaliação documental (*appraisal*), aquela tarefa ingrata e inevitável de escolher o que vai para a reciclagem e o que fica para sempre, decisão que envolve pesar valores, antecipar necessidades futuras, arriscar-se a cometer erros que podem custar caro à história. O problema é que esses princípios, por mais fundamentais que sejam para a prática cotidiana dos arquivistas, raramente são examinados à luz da reflexão filosófica que merecem.

O arquivista aplica a proveniência no dia a dia, mas já parou para pensar no que ela pressupõe sobre o mundo, sobre a natureza dos documentos e suas relações? Ele decide as políticas de acesso, mas já enfrentou de frente a tensão entre o direito à informação e o direito à privacidade, i.e., entre a transparência democrática e a proteção de dados sensíveis? Ele autentica documentos digi-

tais usando assinaturas criptográficas e cadeias de custódia, mas tem clareza sobre o que significa, de fato, um documento digital ser "o mesmo" depois de ter sido migrado de um servidor para outro, de um formato para outro, de um sistema operacional para outro, quando todos os seus bits originais já foram substituídos?

Não são questões abstratas, distantes; elas se tornam urgentes justamente nos casos difíceis, quando o manual não dá conta, quando as regras parecem conflitar, quando o arquivista precisa decidir e não tem uma resposta pronta esperando por ele. E a transição para o mundo digital só fez essa urgência aumentar dramaticamente. Porque, no mundo dos papéis, a materialidade do documento oferecia uma certa segurança, uma âncora física. O papel envelhece, amarela, pode até se desfazer com a umidade ou ser comido por traças, mas está ali, ocupando espaço, resistindo ao tempo de uma forma que a gente pode ver e tocar. No mundo digital, os documentos são cadeias de bits voláteis que precisam ser constantemente migradas, reinterpretadas por novos softwares, revalidadas em novas infraestruturas. O que era implícito e podia ser ignorado se torna explícito e incontornável. O que estava resolvido na prática vira problema teórico da noite para o dia. São precisamente essas questões que a filosofia da arquivologia se propõe a articular com rigor. Sem essa articulação, os princípios da disciplina permanecem como dogmas práticos sem fundamentação teórica explícita, vulneráveis à erosão quando os contextos mudam.

Este capítulo é um convite para fazermos essa travessia juntos. Vamos começar pela metafísica, tentando entender que trem é esse chamado "documento de arquivo" e como ele se relaciona com outros documentos para formar conjuntos maiores, que são mais do que a soma das partes. Depois, vamos para a epistemologia, para ver como podemos conhecer o passado através desses vestígios e como sabemos se podemos confiar neles como fontes confiáveis de informação. E vamos terminar na ética, com as perguntas mais incômodas sobre memória, justiça, privacidade e responsabilidade intergeracional. A ideia não é entregar respostas prontas, mas ajudar a formular melhor as perguntas, porque, na filosofia, uma pergunta bem-feita já é meio caminho andado.

2. Metafísica da Arquivologia

Vamos começar pelo básico, pelo ponto de partida de toda a reflexão arquivística. O que é, afinal, um documento de arquivo? Parece uma pergunta ingênua, quase boba, mas não é. Porque um documento de arquivo não pode ser reduzido a um simples objeto físico, a um pedaço de papel com tinta ou a um arquivo digital com bits. Se fosse só isso, qualquer coisa poderia ser um documento, até uma pedra, um graveto, uma nuvem no céu, um animal empalhado em um museu de história natural. O que faz um pedaço de papel com umas marcas de tinta virar um "documento" é alguma coisa a mais, uma camada extra de significado, de intencionalidade, de função. E é justamente essa camada extra que a metafísica tenta descrever e analisar.

Uma maneira produtiva de começar a desembolar essa questão é contrastar o documento de arquivo com outros tipos de objetos que também carregam informação, mas de um jeito diferente. Pega um livro, por exemplo. Um livro da história de Ouro Preto, desses que a gente encontra em qualquer livraria ou biblioteca. Agora pega um documento de arquivo, digamos, uma ata original da Câmara Municipal de Ouro Preto de 1897. Os dois falam sobre a cidade, os dois podem ser preservados em instituições de memória, os dois têm valor cultural e histórico. Mas a relação deles com o mundo, a maneira como eles existem e funcionam, é radicalmente diferente.

O livro, para começar, foi escrito com a intenção explícita de ser publicado, de circular, de ser lido por um monte de gente que o autor nunca vai conhecer pessoalmente. A função dele, desde o momento em que a primeira palavra foi escrita, é comunicar algo para um público amplo e indefinido. É um objeto feito para ser compartilhado, para ser disseminado, para ser comprado e vendido nas livrarias, para estar disponível em bibliotecas públicas para quem tiver interesse. O autor sabe, desde o início, que está produzindo um objeto que vai circular, que vai ser criticado, que vai ser interpretado de maneiras que ele não pode controlar. Já a ata da Câmara Municipal não foi escrita pensando nos historiadores do futuro, nem no público em geral, nem em quem quer que fosse além dos próprios participantes da reunião e talvez de alguns funcionários da administração encarregados de arquivar os papéis. Ela foi escrita para registrar uma reunião que já estava acontecendo, para documentar uma decisão que estava

sendo tomada naquele exato momento, para servir de prova, se necessário, do que foi discutido e aprovado. A função dela, na sua origem, é prática, imediata, interna à administração da cidade. Que um dia, décadas ou séculos depois, um pesquisador venha a consultá-la é uma possibilidade remota, um acréscimo posterior, não o seu propósito original.

Essa diferença na intencionalidade de criação não é um detalhe psicológico irrelevante. Ela tem consequências ontológicas profundas, que moldam o tipo de entidade que cada um desses objetos é. O livro, como obra, pode existir em milhares de cópias fisicamente distintas. A obra *Dom Casmurro* não é este ou aquele exemplar específico que está na sua estante; é o texto, a estrutura narrativa, a sequência de palavras, que pode ser instanciada em milhares de livros diferentes espalhados pelo mundo. Se um exemplar se perde num incêndio, paciência, a obra continua viva nos outros. Se a editora faz uma nova tiragem, mais mil exemplares vêm ao mundo. Os exemplares são, em grande medida, intercambiáveis para a maioria dos propósitos de leitura. Qualquer exemplar serve para ler o romance, para citar seus trechos, para estudar sua estrutura.

Essa intercambialidade dos exemplares corresponde, nos termos da ontologia analítica contemporânea, à distinção entre tipo (*type*) e instância (*token*), pois a obra é o tipo, os exemplares físicos são as instâncias; a destruição de uma instância não destrói o tipo. O documento de arquivo, ao contrário, não possui essa estrutura tipo-instância no mesmo sentido, pois é, por definição, um testemunho único de um ato único. Assim, a ata original daquela reunião específica da Câmara Municipal é única. Pode haver fotocópias, microfilmes, digitalizações, transcrições feitas à mão por pesquisadores, mas todas são derivadas, cópias, representações, não o original. O original tem um status especial, ontológico e muitas vezes legal, que as cópias não têm. É como uma certidão de nascimento; a fotocópia autenticada pode servir para muitas coisas no dia a dia, para matricular o filho na escola, para tirar passaporte, para provar a idade num concurso, mas se a questão for realmente séria, se houver uma disputa judicial sobre a filiação de alguém, se for preciso provar que a certidão é falsa ou verdadeira, é o original que vai ser exigido em juízo. A unicidade do documento de arquivo é constitutiva do que ele é. Um documento de arquivo é, por definição, um testemunho único de um ato único. Além disso, o documento de arquivo é, por natureza, uma entidade relacional, i.e., sua identidade e sua força probatória

dependem constitutivamente do contexto de criação e das relações com os demais documentos do fundo.

Outra diferença fundamental, que já começamos a explorar, é a relação com o contexto. Um livro, a gente consegue ler e entender razoavelmente bem sem saber tudo sobre a vida do autor ou as circunstâncias políticas e sociais da época em que foi escrito. Claro que o contexto pode enriquecer a leitura, pode adicionar camadas de significado, pode ajudar a entender certas passagens, certas escolhas de palavras, certas referências implícitas. Mas não é condição para a inteligibilidade básica do texto. Uma pessoa no Japão pode ler *Dom Casmurro* numa tradução e entender a história, se emocionar com Capitu, refletir sobre ciúme e traição, sem saber nada sobre o Rio de Janeiro do século XIX, sobre a estrutura da sociedade imperial, sobre a vida de Machado de Assis. O texto, em grande medida, carrega em si mesmo os recursos para sua própria compreensão. Já um documento de arquivo, quando isolado do seu contexto de origem, se torna opaco, muitas vezes ininteligível. Uma ata sem a lista de quem estava presente na reunião, sem a data, sem a indicação de qual órgão a produziu e com que atribuições legais, sem o conhecimento das normas que regiam aquela reunião, sem a compreensão do processo administrativo em que se insere, perde o sentido. Ela pode até ser lida como texto, pode-se decifrar as palavras, mas o seu significado como documento, a sua capacidade de provar algo, a sua função como testemunho de um ato, fica severamente comprometida. O documento de arquivo só volta a fazer pleno sentido quando a gente consegue inseri-lo na teia de relações que o originou, quando a gente consegue conectá-lo aos outros documentos que com ele formam um conjunto, quando a gente entende o lugar que ele ocupava no fluxo de trabalho da entidade que o produziu. Ele é, por natureza, um ser relacional, dependente do contexto para existir como tal.

Tem ainda a questão da função, que já mencionamos de passagem, mas merece ser aprofundada com cuidado, porque é central para tudo o mais. O livro, a gente disse, informa. Ele transmite conhecimento, conta uma história, expõe uma ideia, defende uma tese. O documento de arquivo também informa, claro. A ata informa sobre o que foi discutido e decidido naquela reunião, o ofício informa sobre o que foi solicitado, o relatório informa sobre o que foi feito, a carta informa sobre o que o remetente pensava ou sentia. Mas a sua função

primária, aquilo que o define essencialmente, o seu propósito fundamental, não é meramente informar. É provar. O documento de arquivo é, antes de tudo, um instrumento de prova, de atestação, de testemunho. O que se busca nele, o que se espera dele, o que faz com que ele seja preservado com tanto cuidado, é a autenticidade, a certeza de que aquilo que está registrado corresponde ao que de fato se passou no mundo. É por isso que a gente fala em "fé pública" dos tabeliães e escrivães, em "valor probatório" dos documentos, em "autenticidade" como a qualidade suprema que um documento pode ter. Um livro pode ser uma ficção, pode ser um romance, pode ser uma obra de imaginação, e isso não o desqualifica como livro; pelo contrário, pode ser exatamente isso que o torna valioso, que o torna arte. Um documento de arquivo que não seja autêntico, que seja uma falsificação ou uma adulteração, perde a sua razão de ser. Ele pode até continuar existindo como objeto físico, pode até ser estudado como exemplo de falsificação, como evidência de uma tentativa de fraude, mas deixa de ser, propriamente, um documento de arquivo no sentido pleno da palavra.

Essas características todas, a origem orgânica e não intencionalmente pública, a unicidade, a dependência constitutiva de contexto e a função primordialmente probatória, já foram estudadas com um rigor impressionante por uma disciplina que parece coisa de outro século, uma relíquia do passado, mas que voltou com toda a força na era digital, a saber, a diplomática (*diplomatics*).

A diplomática nasceu no século XVII, num contexto muito específico e turbulento da história europeia. Havia uma enxurrada de documentos medievais circulando pela Europa, e muitos deles eram falsificações. Mosteiros falsificavam documentos para provar que tinham direitos sobre terras, reis falsificavam para justificar a cobrança de impostos, particulares falsificavam para ganhar causas na justiça, para provar linhagens nobres que não tinham, para reivindicar heranças. O monge beneditino Jean Mabillon (1681), diante dessa confusão generalizada, diante da dificuldade de distinguir o verdadeiro do falso num mar de papéis, desenvolveu um método rigoroso para separar o joio do trigo, para distinguir os documentos genuínos dos falsos com algum grau de segurança. Ele examinava o tipo de pergaminho usado, a composição da tinta, o estilo da caligrafia, os formulários textuais típicos de cada chancelaria real ou eclesiástica, os selos e assinaturas, a presença de anacronismos, as marcas de uso e de desgaste. A ideia central era que a autenticidade se revelava na materialidade,

na forma, no contexto, nos mínimos detalhes que um falsificador descuidado ou apressado poderia negligenciar. O documento era dissecado como um corpo numa autópsia, e cada detalhe podia ser uma pista a favor ou contra a sua pretensão de ser genuíno.

A grande arquivista contemporânea Duranti (1998) percebeu, já nos anos 1990, quando o mundo digital começava a se impor como realidade incontornável, que essa velha diplomática, desenvolvida para lidar com pergaminhos e selos de cera, com caligrafias e tintas ferrogálicas, era na verdade a chave para entender os desafios colocados pelos documentos digitais. Porque, por baixo da diferença colossal de suporte (pergaminho e bits, selo de cera e assinatura digital, caligrafia e fonte tipográfica), os problemas fundamentais são os mesmos. Como saber se um documento é mesmo quem ele diz ser? Como garantir que não foi adulterado ao longo do tempo, que não sofreu alterações não autorizadas? Como assegurar que a sua forma atual corresponde à forma que tinha quando foi criado e quando era válido? A diplomática, pensada para o mundo medieval, ganhou assim uma nova vida no século XXI, tornando-se uma ferramenta conceitual indispensável para arquivistas que lidam com documentos nativos digitais, com e-mails, com bancos de dados, com sistemas de informação complexos.

Para a diplomática, um documento não é uma coisa simples, de uma só camada, transparente e fácil de entender. Ele tem uma anatomia, uma estrutura complexa que a gente pode desmontar pra entender como cada parte contribui para a sua função e para a sua autenticidade. Vamos pensar nisso com calma, porque é central para toda a reflexão que vem a seguir.

Tudo começa com o contexto de criação. Não dá pra entender um documento, não dá pra avaliar a sua autenticidade, não dá pra usar o seu conteúdo como prova, sem saber de onde ele veio, quem o fez, por que foi feito, em que circunstâncias, com que autoridade. É como encontrar uma carta antiga dentro de um livro; você imediatamente quer saber quem escreveu, para quem, em que data, em que cidade, qual era a relação entre eles, qual era o assunto. Essas informações não são enfeites, não são detalhes supérfluos para curiosos; elas são parte do que a carta é. No caso dos documentos de arquivo, esse contexto tem várias camadas que se entrelaçam e se sobrepõem. Tem o contexto jurídico-administrativo, que são as leis, os regulamentos, as normas, os procedimen-

tos que estavam em vigor no momento em que o documento foi criado e que davam a ele a sua força, a sua validade, a sua autoridade. Uma ata de reunião, para ser válida como documento oficial, para ter valor probatório, precisa seguir certas regras estabelecidas em lei ou em regimentos internos. Precisa ter uma lista de presentes, precisa registrar as deliberações, precisa ser assinada por quem de direito. Se não seguiu essas regras, pode ser que ela não valha nada como prova, pode ser que seja juridicamente nula, mesmo que o conteúdo factual esteja perfeitamente correto. Tem o contexto de proveniência, que é a entidade que produziu o documento, seja uma pessoa, uma família, uma empresa, um órgão público, uma associação. Saber se a ata veio da Secretaria de Saúde ou da Secretaria de Cultura não é um detalhe irrelevante; muda completamente o sentido do que está registrado, a autoridade daquele registro, o seu peso como prova sobre determinados assuntos. Uma decisão da Secretaria de Saúde sobre um hospital não tem o mesmo valor que uma decisão da Secretaria de Cultura sobre o mesmo hospital. Tem ainda o contexto procedimental, que é o processo maior em que aquele documento se insere, a sequência de ações de que ele faz parte. Um ofício nunca vem sozinho, isolado no mundo; ele é resposta a um ofício anterior, faz parte de uma troca, de um diálogo, de um fluxo de trabalho, e vai gerar uma cadeia de outros documentos, ofícios de resposta, despachos internos, anexos, pareceres. E tem o contexto documental, as relações com outros papéis específicos, como as referências explícitas, as citações, os anexos, as minutas, as versões preliminares, as cópias. Cada documento é um ponto numa teia, e o seu significado, a sua força probatória, a sua própria inteligibilidade dependem em grande parte da sua posição nessa teia.

Depois vem a forma documental. Aqui a gente olha pro documento como um objeto, com sua materialidade e sua estrutura, com sua aparência e sua organização interna. No mundo físico, a forma material inclui o tipo de suporte, se é papel comum, papel timbrado, pergaminho, se tem marcas d'água, o tipo de tinta usada, a presença de selos, carimbos, assinaturas, a disposição espacial dos elementos na página, as dobras, os rasgos, as marcas de uso. No mundo digital, os equivalentes funcionais seriam o formato do arquivo, os metadados técnicos embutidos, as assinaturas digitais, os hashes criptográficos, os certificados, as marcas d'água digitais, os registros de log. Tudo isso carrega informação sobre autenticidade, sobre a história do documento, sobre as suas

transformações. Mas a forma não é apenas material; é também intelectual, isto é, a estrutura sintática do texto, a maneira como as informações são organizadas e apresentadas, as convenções de estilo e de formato que o documento segue. Uma ata tem uma estrutura típica, um formato reconhecível por qualquer pessoa familiarizada com administração; ela começa com a identificação da reunião, a data, o local, os participantes presentes e ausentes, depois vem a pauta, o registro das discussões, as deliberações e votações, e por fim o encerramento com a data e as assinaturas. Um contrato tem outra estrutura, com a qualificação das partes, o objeto do contrato, as cláusulas, o foro, as testemunhas. Uma petição tem outra, com o endereçamento à autoridade, a exposição dos fatos, os fundamentos jurídicos, o pedido. Essas estruturas não são arbitrárias, não são invenção de burocratas entediados sem nada melhor para fazer; elas são padronizadas justamente para que o documento possa cumprir sua função de forma confiável, para que qualquer pessoa minimamente familiarizada com aquele tipo de documento saiba onde encontrar cada informação e o que esperar de cada parte. A forma documental é um dispositivo de previsibilidade e confiança, uma maneira de garantir que o documento seja interpretado corretamente por aqueles a quem se destina.

Por fim, tem o conteúdo, que é o que o documento diz, o seu texto propriamente dito, as palavras que o compõem. Mas a diplomática nos ensina a olhar para o conteúdo também de forma estruturada, não como uma massa amorfa de palavras, mas como algo que tem uma lógica interna, uma arquitetura própria. O conteúdo articula a ação que o documento executa ou registra. Por exemplo, um decreto nomeia alguém para um cargo, exonera outro, cria uma comissão; um contrato estabelece obrigações recíprocas entre as partes; uma ata certifica que uma decisão foi tomada, que uma votação ocorreu, que um resultado foi proclamado. Ele especifica as pessoas envolvidas (o autor, o destinatário, as testemunhas, os interessados, outros mencionados) e os papéis que cada uma desempenha naquele ato. Ele trata do assunto específico, usando muitas vezes uma linguagem técnica, cheia de termos próprios do direito, da administração, do comércio, da ciência, termos que precisam ser compreendidos no seu contexto para que o sentido do documento seja apreendido. E ele ancora o documento no tempo e no espaço, através de datas e lugares de

emissão, que são informações indexicais fundamentais para situar o ato documentado na história.

O que a análise diplomática revela, com toda a clareza que a gente poderia desejar, é que o documento de arquivo não é um mero portador neutro de informação, um envelope vazio que poderia conter qualquer mensagem indiferentemente. Ele é uma entidade complexa e estruturada, cujas diferentes camadas (material, formal, contextual, de conteúdo) estão intrinsecamente ligadas à sua função probatória. A forma não é um invólucro acidental para o conteúdo, algo que se poderia mudar sem prejuízo; ela é parte constitutiva do que faz do documento um documento autêntico e confiável. Alterar a forma, desrespeitar a estrutura esperada, omitir elementos contextuais, não é apenas uma questão de apresentação, de estética, de organização; é comprometer a própria capacidade do documento de cumprir a sua função essencial.

Essa complexidade aponta para uma questão ontológica profunda, que tem ocupado filósofos e arquivistas há décadas e que se torna cada vez mais premente na era digital. O documento de arquivo parece ter uma natureza dual. Por um lado, ele é inegavelmente um objeto físico, com existência material no mundo, ocupando espaço, tendo massa, estando sujeito às leis da física e da química. O papel, a tinta, o selo de cera, o disco rígido onde os bits estão gravados, a fita magnética, o CD-ROM, tudo isso é matéria, deteriora-se com o tempo, pode ser queimado, molhado, rasgado, apagado. Por outro lado, o documento é também uma entidade lógica, intelectual, que carrega um conteúdo semântico, que se refere a pessoas, eventos, lugares no mundo, que pode ser verdadeiro ou falso, que pode ser copiado e transferido para outros suportes sem perder a sua identidade funcional. A mesma ata, o mesmo conteúdo proposicional, a mesma sequência de palavras com o mesmo significado, pode existir no papel original, numa fotocópia, num microfilme, num arquivo PDF, num printout, numa transcrição feita à mão. Em que sentido é “a mesma” ata? O que faz com que todas essas manifestações materiais diferentes sejam manifestações da mesma coisa?

Essa dualidade não é mero quebra-cabeça filosófico, pois tem consequências práticas de primeira ordem que afetam decisões concretas de preservação, restauração, autenticação e acesso. Pense num documento que passou por um processo de restauração. O papel original estava se desfazendo, amarelado,

frágil, prestes a virar pó. O restaurador, com todo o cuidado e técnica, substituiu as partes danificadas por papel novo, especialmente preparado para se assemelhar ao original em cor, textura e espessura. A tinta original foi consolidada, mas em alguns trechos onde estava quase apagada teve que ser reforçada com tinta nova, aplicada com precisão milimétrica. No final do processo, a matéria física do documento já não é a mesma de antes; grande parte do que está ali agora é um material novo, introduzido pela restauração. No entanto, o documento restaurado é apresentado ao público, nos museus e arquivos, como o mesmo documento histórico que era antes, a mesma carta, a mesma ata, o mesmo pergaminho. Até que ponto isso é verdade? Se a identidade do documento estivesse na sua materialidade original, naqueles átomos específicos que o compunham, ele teria deixado de existir no momento em que o primeiro pedaço de papel foi substituído. Mas a prática museológica e arquivística, e a nossa própria intuição, tratam o documento como persistente através dessas transformações materiais. Algo permanece, algo que não é redutível à matéria.

O mesmo problema, em versão ainda mais radical e desafiadora, aparece nos documentos digitais. Um arquivo digital é, por natureza, copiável e migrável sem perda aparente. Para preservá-lo a longo prazo, para garantir que ele sobreviva à obsolescência tecnológica, é necessário, periodicamente, copiá-lo para novas mídias quando o CD ou o HD antigo começa a degradar-se ou quando o dispositivo para lê-los se torna obsoleto. Nesse sentido, é necessário, muitas vezes, convertê-lo para novos formatos. Depois de algumas décadas e algumas migrações, depois de algumas cópias e algumas conversões, não resta um único bit físico original. Tudo foi substituído, cada 0 e cada 1 foram reescritos em novos suportes, em novos formatos. No entanto, o arquivo digital é considerado o mesmo documento, a mesma fotografia, o mesmo texto, o mesmo vídeo, desde que o processo de migração tenha sido feito corretamente, preservando o conteúdo e as propriedades funcionais essenciais. A identidade do documento digital parece residir não na sua materialidade física, que é volátil, efêmera e constantemente substituída, mas na sua continuidade lógica, na sua história de migrações, na cadeia de custódia que atesta a correção desses processos.

O que essas reflexões sugerem, de forma cada vez mais clara, é que o documento de arquivo não pode ser identificado simplesmente com o seu suporte

material, com a matéria de que é feito num dado momento. A relação entre o documento e o objeto físico que o corporifica num dado instante é antes uma relação de constituição, no sentido técnico que a filósofa Lynne Rudder Baker (2007) deu ao termo. O objeto físico, o papel com tinta ou o conjunto de bits no disco rígido, constitui o documento naquele momento, mas não é estritamente idêntico a ele. O documento pode persistir através de mudanças no objeto que o constitui, pode ser constituído por diferentes objetos em diferentes momentos, desde que haja uma continuidade causal e funcional apropriada, e desde que as práticas institucionais, como a documentação da cadeia de custódia e dos procedimentos de migração, sustentem e atestem essa persistência.

Essa análise ontológica, um pouco abstrata talvez, mas necessária, ajuda a entender por que princípios como o da proveniência e o da ordem original são tão centrais para a arquivologia, por que eles são ensinados em todos os cursos, repetidos em todos os manuais, respeitados em todas as instituições. Eles não são meras convenções organizacionais, inventadas para facilitar a vida dos arquivistas, para padronizar procedimentos, para tornar o trabalho mais eficiente. Eles são o reconhecimento, talvez implícito, talvez não totalmente articulado, de que a identidade e o valor dos documentos são, em parte, constituídos pelas suas relações contextuais. Um documento isolado, separado do seu fundo de origem, desprovido da sua ordem original, sem conexão com os outros documentos que com ele formavam um conjunto, é um documento mutilado, que perdeu parte do seu ser, parte da sua capacidade de significar, parte da sua força como prova. O fundo arquivístico, por sua vez, não é uma coleção arbitrária de itens justapostos por conveniência, mas um todo orgânico, uma entidade com identidade própria, cujas partes se iluminam mutuamente, cujas relações são tão importantes quanto os próprios documentos. A proveniência e a ordem original são, portanto, princípios ontológicos, não apenas regras de arrumação.

Essa visão organicista do fundo, no entanto, não está imune a objeções, e é importante considerá-las com cuidado para não cair num realismo ingênuo. Alguém poderia argumentar, de uma perspectiva mais reducionista ou nominalista, que o fundo não passa de uma coleção de objetos físicos agregados por conveniência humana, por decisões de arquivistas, por práticas institucionais. Não haveria ali nenhuma propriedade emergente genuína, nada que fosse mais do que a soma das partes; tudo o que existe são documentos individuais, e as

relações entre eles são relações que os arquivistas projetam sobre os objetos, não algo que esteja lá no mundo para ser descoberto independentemente de nós. Essa crítica ecoa alguns debates mais amplos e antigos na metafísica sobre a relação entre partes e todos, sobre se um todo pode ser algo mais do que a mera soma de suas partes constituintes, e sobre se as relações são reais ou apenas construções mentais.

A resposta a essa objeção, me parece, passa por distinguir entre diferentes tipos de propriedades e diferentes tipos de relações. O peso total de um fundo, por exemplo, é uma propriedade meramente agregativa, pois é a soma dos pesos de cada documento, e nada mais. Já a capacidade probatória do fundo, a sua coerência narrativa, a sua integridade como testemunho das atividades de uma entidade ao longo do tempo, não são propriedades que se deixam reduzir dessa forma simples. Elas são propriedades que emergem justamente das relações entre os documentos, das conexões funcionais e contextuais que os unem, da maneira como se iluminam mutuamente. Um único ofício, isolado de seu contexto, pode não provar grande coisa, pode ser ambíguo, pode ser mal interpretado. Mas esse mesmo ofício, quando reinserido na sequência documental a que pertence, quando lido em conjunto com os ofícios que o precederam e o sucederam, quando compreendido à luz do processo administrativo de que fazia parte, dentro do seu fundo de origem, passa a ser uma peça fundamental de um quebra-cabeça maior, uma evidência robusta de uma decisão, de um fluxo de trabalho, de uma política, de uma mudança institucional.

O princípio da ordem original, que é um corolário natural do princípio da proveniência, reforça essa ontologia relacional. A ordem original não é uma arrumação qualquer que o arquivista inventa para organizar o material, uma classificação subjetiva entre outras possíveis. É a disposição que os documentos tinham quando ainda eram ativamente usados pelo seu criador, quando faziam parte do dia a dia da entidade que os produziu, quando estavam vivos, por assim dizer, nos fluxos de trabalho e nos processos de decisão. Essa ordem é, ela mesma, informativa. Ela revela fluxos de trabalho, prioridades administrativas, hierarquias funcionais, maneiras de classificar e arquivar que eram próprias daquela entidade, daquela época, daquele contexto. Ela nos diz como as coisas eram feitas, o que era importante, o que vinha antes, o que vinha depois. Violar a ordem original, reorganizar os documentos segundo um critério externo

qualquer, por mais lógico que pareça, é perder uma camada inteira de informação contextual que é constitutiva do significado dos documentos, é apagar vestígios preciosos do funcionamento passado.

A aplicação do princípio da ordem original em contextos digitais, no entanto, torna-se filosoficamente espinhosa e exige criatividade conceitual. Documentos digitais frequentemente não possuem uma ordem original no mesmo sentido espacial ou físico que os documentos em papel. Eles existem em diretórios que podem ser reorganizados com um clique do mouse, em sistemas de arquivos onde a ordem é arbitrária, em bancos de dados onde a ordem de apresentação é dinâmica e dependente de consultas, em sistemas de gestão onde a mesma informação pode ser acessada por diferentes caminhos, em nuvens onde a localização física dos bits é irrelevante e desconhecida. O que significa, então, preservar a ordem original de um conjunto de e-mails, de arquivos em um servidor, de registros em um sistema de gestão documental, de posts em uma rede social? Essa dificuldade prática, que os arquivistas enfrentam diariamente, revela um problema metafísico mais profundo, o de que a ordem original, como concebida para o mundo físico dos papéis e das pastas, pressupõe uma materialidade fixa e uma espacialidade estável que os objetos digitais não possuem necessariamente. Talvez seja necessário reinterpretar a ordem original não como um arranjo físico, mas como uma estrutura lógica ou contextual. O que importa preservar não é a localização física dos bits, que é efêmera e irrelevante para a identidade do documento, mas sim as relações funcionais, temporais e processuais entre os documentos digitais, a maneira como eles se conectam, se referem uns aos outros, se sucedem, se implicam. Metadados que registrem a data de criação, o autor, o destinatário, as relações de resposta ou de versão, os links, as referências, podem capturar a ordem original em um sentido relacional, não espacial, e podem reconstruir a teia de conexões que dava sentido aos documentos no seu contexto de uso ativo. Essa reinterpretação, no entanto, exige que repensemos o que constitui a identidade e a integridade de um fundo em ambiente digital, menos como um conjunto de objetos fixos em estantes, mais como uma rede de entidades interconectadas cujas relações são parte de sua essência e cuja preservação exige a preservação dessas relações, não apenas dos objetos isolados.

3. Epistemologia da Arquivologia

Depois de toda essa discussão sobre o que os documentos são, sobre a sua natureza ontológica complexa e relacional, chega a hora de fazer uma pergunta diferente, mas igualmente fundamental, e talvez ainda mais urgente do ponto de vista prático, a saber, como a gente conhece através deles? Que tipo de conhecimento os arquivos podem nos proporcionar sobre o passado, sobre as pessoas, sobre as instituições e sobre os eventos? E, mais importante ainda, como sabemos que podemos confiar neles? Como distinguimos um documento autêntico, que merece fé, que pode ser usado como prova, de uma falsificação, de uma adulteração, de um documento que parece ser o que não é?

Essa pergunta pela confiabilidade epistêmica dos documentos articula-se com o problema mais amplo da justificação do testemunho (*testimony*) na epistemologia contemporânea, de saber em que condições é racional confiar no que outros nos transmitem, quando não podemos verificar diretamente a fonte. Os arquivos são formas institucionalizadas de testemunho mediado, e sua epistemologia precisa responder ao mesmo desafio.

Essa é a questão epistemológica central da arquivologia, o seu problema do conhecimento por excelência. Porque os arquivos não são apenas depósitos de papel velho, de documentos empoeirados, de registros esquecidos; eles são, ou pelo menos deveriam ser, fontes de conhecimento sobre o passado, base para a escrita da história, fundamento para direitos, instrumentos de transparência e de prestação de contas. Historiadores, jornalistas, cidadãos comuns, advogados, juízes, gestores públicos, todos recorrem aos arquivos quando precisam saber o que realmente aconteceu, quando precisam de provas para fundamentar suas alegações, quando precisam basear suas decisões em algo mais sólido do que a memória falível, o boato ou a opinião interessada. A pluralidade de usuários aqui enumerada é relevante para a epistemologia arquivística porque implica que os arquivos precisam satisfazer critérios de confiabilidade distintos segundo contextos distintos. A prova jurídica, a evidência histórica, o dado administrativo e a fonte jornalística operam sob regimes epistêmicos diferentes, com diferentes limiares de aceitação, diferentes modalidades de verificação e diferentes consequências para o erro. Mas essa confiança nos arquivos, essa crença de que eles nos dão acesso ao passado, só é justificada se houver ma-

neiras confiáveis de garantir que os documentos são o que pretendem ser e que a informação que contêm chegou até nós sem distorções inaceitáveis ao longo do tempo.

O conceito de autenticidade (*authenticity*) é, por isso, absolutamente central na epistemologia arquivística. É o conceito em torno do qual gira toda a confiança que depositamos nos documentos. Um documento autêntico é aquele que é genuíno, que é o que pretende ser, que foi criado por quem diz ter sido criado, no momento em que diz ter sido criado, no contexto em que diz ter sido criado, e que não sofreu alterações não autorizadas, adulterações, falsificações, desde então. Mas essa definição, que parece simples e direta à primeira vista, esconde uma série de complexidades filosóficas que precisam ser desembrulhadas com cuidado, sob pena de tomarmos a autenticidade como algo mais simples do que realmente é.

Na tradição diplomática, a autenticidade foi historicamente definida a partir de critérios formais, presença de selos, subscrições, fórmulas protocolares, que constituíam a forma documental como dispositivo de confiança. A modernidade desloca gradualmente o critério formal para critérios contextuais e processuais, tornando a autenticidade dependente da cadeia de custódia e do controle institucional, não apenas da forma do documento.

Nesse sentido, o que significa, exatamente, dizer que um documento "é o que pretende ser"? Essa formulação já sugere que a autenticidade não é uma propriedade puramente intrínseca do documento, algo que se possa ler diretamente na sua materialidade, como se lê a cor de um objeto ou se mede o seu peso. A autenticidade envolve uma relação entre o documento e uma pretensão, uma alegação que ele faz sobre si mesmo, uma identidade que ele reivindica. O documento se apresenta como sendo algo; por exemplo, como sendo uma ata da Câmara Municipal de 1897, como sendo uma carta de Getúlio Vargas, como sendo um decreto de Dom Pedro I. A autenticidade é a propriedade de essa pretensão ser verdadeira, de essa alegação corresponder aos fatos. Mas como saber se a pretensão é verdadeira? Como verificar se o documento é mesmo o que diz ser?

A caracterização da autenticidade como relação entre documento e pretensão aproxima-se do conceito de intencionalidade desenvolvido na fenomenologia husserliana, pois o documento não é um objeto inerte, mas um objeto inten-

cional, ele é sobre algo, refere-se a algo fora de si mesmo, e sua autenticidade consiste precisamente na correspondência entre essa referência e aquilo a que pretende referir-se. Isso implica que a autenticação é sempre um ato interpretativo, não um ato de leitura direta de propriedades físicas.

Essa verificação, tradicionalmente, dependia de uma combinação de fatores físicos; porém dependia também da confiança nas instituições que o custodiaram ao longo do tempo, na cadeia de custódia que atesta a sua integridade. A autenticidade, portanto, não é algo que se lê diretamente no documento; é algo que se infere a partir de múltiplas evidências, que se constrói através de um raciocínio complexo que envolve diferentes tipos de conhecimento. Essa natureza inferencial da autenticidade tem uma consequência epistemológica importante, a saber, ela implica que a autenticação é sempre falível, sempre revisável. Como qualquer inferência a partir de evidências, ela está sujeita a erro, pode ser contestada por novas evidências, pode ser revertida. Não há autenticação definitiva e irrevogável; há graus de confirmação que aumentam ou diminuem conforme as evidências disponíveis. Isso não é uma fraqueza da arquivologia, mas uma característica estrutural do conhecimento baseado em evidências.

A cadeia de custódia (*chain of custody*) é também um conceito epistemológico fundamental aqui. Para que possamos confiar que um documento chegou até nós sem adulterações, precisamos de um registro contínuo e ininterrupto de todos os seus custodiantes, de todas as instituições e pessoas que o guardaram, de todas as transferências de posse, de todas as condições de armazenamento, de todos os procedimentos de conservação a que foi submetido. Essa cadeia funciona como um rastro, um vestígio, uma testemunha da história do documento depois da sua criação. Se a cadeia é quebrada, se há lacunas, se há períodos em que não se sabe onde o documento esteve ou em que mãos, a autenticidade fica comprometida, porque não podemos garantir que nesses períodos de sombra não ocorreram adulterações, substituições, danos. A cadeia de custódia é, assim, um dispositivo epistemológico que estende a nossa confiança no documento para além do que podemos verificar diretamente, ancorando-a na confiança que depositamos nas instituições e nos procedimentos que documentaram a sua história.

A analogia com o conceito jurídico de cadeia probatória não é acidental; em ambos os casos, a admissibilidade epistêmica de um item de evidência depende da demonstração de que sua integridade foi mantida ao longo de todo o percurso desde sua origem até o momento de uso. A diferença é que a cadeia de custódia arquivística opera em escalas temporais imensamente maiores, o que amplifica tanto a importância quanto a fragilidade do dispositivo.

No mundo digital, esse conceito se torna ainda mais central e, ao mesmo tempo, mais complexo. Porque no mundo digital não há um "original" material estável que possa ser guardado num cofre. O que há são cópias, migrações, backups, replicações. A cadeia de custódia digital precisa registrar não apenas quem custodiou o quê, mas também quando e como os dados foram copiados, migrados, convertidos, verificados. Precisa incluir informações técnicas sobre formatos, sobre hashes, sobre assinaturas digitais, sobre metadados de preservação, além de garantir que esses registros da cadeia sejam eles mesmos preservados e autênticos, sob pena de entrarmos num regresso infinito.

O regresso infinito aqui mencionado não é meramente retórico. Ele nomeia um problema estrutural da preservação digital que a literatura especializada denomina problema da autenticidade de segunda ordem (*second-order authenticity)*. Para garantir a autenticidade dos documentos digitais, produzem-se metadados de preservação; mas esses metadados são eles próprios documentos digitais cuja autenticidade precisa ser garantida; e os registros que garantem sua autenticidade são, por sua vez, documentos digitais. Infelizmente, não temos uma solução lógica para esse problema, mas somente soluções práticas e institucionais, como ter convenções interoperáveis, auditorias externas, padrões compartilhados.

Há também o problema da confiança nos instrumentos de autenticação. No mundo digital, a autenticidade é frequentemente garantida por mecanismos criptográficos, como assinaturas digitais, hashes, blockchain. Mas esses mecanismos dependem, eles mesmos, de infraestruturas de confiança. Uma assinatura digital só é válida se confiarmos na autoridade certificadora que a emitiu. Um hash só garante a integridade se confiarmos que o algoritmo não foi quebrado e que o hash foi corretamente calculado no momento certo. A blockchain só é confiável se confiarmos no protocolo e na rede que a mantém. A autenticidade digital, portanto, não elimina a necessidade de confiança; apenas a deslo-

ca para outros pontos, para as infraestruturas técnicas e institucionais que sustentam os mecanismos de autenticação.

Esse deslocamento da confiança é uma das contribuições mais importantes da arquivologia digital para a epistemologia geral, pois demonstra que nenhum mecanismo técnico de garantia de autenticidade é autossuficiente, toda cadeia de verificação pressupõe, em algum ponto, uma ancoragem em uma confiança institucional que não pode ser reduzida a procedimento técnico. A criptografia garante a integridade matemática, mas não a confiança social.

Isso nos leva a uma questão epistemológica mais profunda, a de que a confiança nos arquivos não é uma questão puramente técnica. Ela é também, e talvez fundamentalmente, uma questão social e institucional. Confiamos nos arquivos porque confiamos nas instituições que os mantêm, nos profissionais que os organizam, nas práticas que os regem. Confiamos na competência dos arquivistas, na sua honestidade, no seu compromisso com a verdade. Confiamos que eles não vão adulterar documentos, que não vão destruir provas inconvenientes, que não vão favorecer certas interpretações em detrimento de outras. Essa confiança institucional é frágil, pode ser traída, precisa ser constantemente renovada e justificada, mas é indispensável.

Na epistemologia social, esse tipo de confiança institucional é estudado sob o conceito de autoridade epistêmica legítima, que seria a atribuição de credibilidade a uma fonte não com base no exame direto de cada afirmação que ela produz, mas com base em avaliações de segunda ordem sobre sua competência, honestidade e filiação a práticas epistêmicas estabelecidas. A confiança nos arquivos é, nesse sentido, constitutivamente, uma forma de confiança epistêmica de segunda ordem.

Outra questão epistemológica fundamental diz respeito à relação entre o documento individual e o conjunto mais amplo em que ele se insere. Um documento isolado pode ser enganoso. Pode registrar apenas um lado de uma história, pode omitir informações cruciais, pode refletir os preconceitos do seu autor, pode ser um caso atípico. A força probatória de um documento aumenta enormemente quando ele é confrontado com outros documentos, quando é inserido em séries, quando pode ser corroborado por fontes independentes. É aí que entra a importância do fundo e da ordem original para a epistemologia arquivística. A organização dos documentos em fundos, a preservação das suas rela-

ções originais, não é apenas uma questão de arrumação; é um dispositivo epistemológico que nos permite fazer inferências mais robustas, que nos permite cruzar informações, que nos permite contextualizar e interpretar corretamente cada peça individual.

A pertença de um documento a um fundo é, portanto, uma qualidade epistêmica, não apenas administrativa. Ela determina o conjunto de outros documentos em relação aos quais ele pode ser interpretado, e transforma cada documento num nó de uma rede de relações que constitui o contexto de sua inteligibilidade. Retirar um documento de seu fundo é empobrecer epistemicamente sua capacidade probatória, o que explica por que a dispersão de arquivos em contextos coloniais representa não apenas uma perda patrimonial, mas uma perda epistêmica.

Um historiador que trabalha com documentos de arquivo sabe que não pode confiar cegamente em nenhum documento isolado. Ele precisa entender o contexto em que o documento foi produzido, a posição do seu autor, os interesses em jogo, as convenções da época. Precisa comparar diferentes documentos sobre o mesmo assunto, confrontar versões, buscar contradições, identificar silêncios. Precisa, em suma, fazer uma crítica das fontes, um trabalho de interpretação que é, ele mesmo, um exercício epistemológico complexo. E esse trabalho é enormemente facilitado quando os documentos são preservados em seus contextos originais, quando as relações entre eles são mantidas, quando a proveniência é respeitada. A ordem original é, de fato, uma ferramenta cognitiva, um mapa que orienta a investigação.

O conceito de silêncio arquivístico (*archival silence*) é aqui relevante. Os documentos que não existem, porque nunca foram criados, porque foram destruídos, porque estão inacessíveis, são também fontes de conhecimento, na medida em que permitem inferir o que foi sistematicamente excluído do registro. A crítica das fontes arquivísticas inclui, portanto, não apenas o exame do que está presente, mas a atenção ao que está ausente e ao que essa ausência pode significar.

4. Ética da Arquivologia

Depois de tantas questões sobre o que os documentos são e sobre como podemos conhecer através deles, chegamos talvez às perguntas mais urgentes e delicadas de todas, aquelas que tocam em feridas abertas, em conflitos não resolvidos, em tensões que não admitem soluções fáceis, as perguntas éticas. Que responsabilidades os arquivistas têm para com os objetos que custodiam, para com as pessoas que criaram esses objetos, para com aqueles que são representados neles, e para com a sociedade como um todo, presente e futura? Como equilibrar valores que muitas vezes entram em rota de colisão, como o direito à informação e o direito à privacidade, a transparência e a proteção de dados sensíveis, a preservação da memória e o direito ao esquecimento, o acesso universal e os direitos das comunidades de origem? Essas questões constituem o que pode ser chamado de ética arquivística aplicada, um domínio em que princípios éticos gerais (autonomia, beneficência, não maleficência, justiça) precisam ser traduzidos em critérios operacionais para decisões concretas sobre aquisição, avaliação, acesso e descrição de documentos. A especificidade dessas questões em relação à ética geral reside no fato de que os arquivos afetam interesses de sujeitos que não participaram de sua criação e que frequentemente não têm voz nos processos de sua gestão.

Uma das questões éticas mais vivas na arquivologia contemporânea, e que tem ganhado cada vez mais espaço nos debates públicos e acadêmicos, diz respeito à propriedade do patrimônio cultural e, mais especificamente, aos pedidos de repatriação de documentos e objetos que foram levados de seus locais de origem, muitas vezes em contextos coloniais, por meios que hoje consideraríamos ilegítimos, imorais ou francamente criminosos. Museus e arquivos europeus e norte-americanos estão cheios de materiais que foram pilhados em guerras, saqueados em expedições, comprados em transações profundamente assimétricas, em que uma das partes não tinha real poder de barganha ou sequer compreendia o que estava sendo negociado, ou simplesmente retirados de comunidades indígenas e de países colonizados sem o devido consentimento, sob a justificativa de que os “nativos” não saberiam preservar adequadamente o seu próprio patrimônio. Os Mármores de Elgin, no Museu Britânico, são talvez o exemplo mais famoso e debatido, mas há inúmeros outros, como os bronzes do

Benin espalhados por museus da Europa e da América, os artefatos das culturas pré-colombianas em coleções particulares e institucionais, os documentos de antigas colônias africanas e asiáticas em arquivos das antigas metrópoles, os objetos sagrados de povos indígenas em vitrines de museus de antropologia, os restos humanos de ancestrais indígenas em gavetas de museus de história natural.

O aspecto arquivístico específico desse problema, a custódia de documentos coloniais em arquivos de antigas metrópoles, acrescenta uma dimensão epistemológica à dimensão ética, dado que esses documentos foram frequentemente produzidos por e para a administração colonial, registrando os povos colonizados a partir da perspectiva do colonizador. Sua repatriação não é apenas uma questão de propriedade, mas de quem controla o registro do passado e, portanto, quem tem autoridade sobre a narrativa histórica.

Diante desse cenário, a pergunta é inevitável e não pode ser ignorada. Os arquivos e museus que hoje custodiam esses materiais têm a obrigação moral de devolvê-los às comunidades ou países de origem? A questão é complexa, admite respostas muito diferentes, e envolve uma série de considerações que precisam ser pesadas com cuidado.

De um lado, há quem defenda uma posição cosmopolita, argumentando que o patrimônio cultural pertence, em última instância, à humanidade como um todo, e não a nações ou comunidades específicas. James Cuno (2008), por exemplo, que dirigiu por anos o *Art Institute of Chicago* e depois o Getty Trust, defende em seus livros que os chamados "museus universais", como o Louvre, o Metropolitan Museum, o Museu Britânico, cumprem uma função moral importante, ao tornar acessível a um público global, diverso e numeroso, obras de arte e documentos de diversas culturas e épocas, promovendo assim a compreensão intercultural, o apreço pela diversidade da produção humana, e o estudo comparado que seria impossível se cada objeto estivesse confinado ao seu país de origem. Repatriar esses objetos, segundo essa visão, seria fragmentar esse patrimônio comum, dispersar coleções que foram formadas ao longo de séculos, e reduzir o acesso global a eles, confinando as obras a seus locais de origem, onde talvez não tenham a mesma infraestrutura de preservação, o mesmo público, a mesma visibilidade internacional.

A posição cosmopolita de Cuno (2008) pressupõe uma ontologia do patrimônio como bem público global que está longe de ser consensual; ela universaliza critérios estéticos e de acesso que são, na prática, os critérios das instituições dos países desenvolvidos, e ignora que o conceito de "patrimônio da humanidade" foi historicamente instrumentalizado para legitimar a retenção de objetos obtidos por meios ilegítimos.

De outro lado, os críticos dessa posição, e eles são muitos e cada vez mais influentes, apontam que o discurso da "humanidade" e do "patrimônio comum" muitas vezes serve para mascarar a continuidade de relações de poder profundamente assimétricas, para justificar a manutenção do *status quo*, para silenciar as vozes dos que foram espoliados. Na prática, as decisões sobre o que preservar, como exibir, a quem dar acesso, que narrativas contar, continuam sendo tomadas por curadores europeus e norte-americanos, a partir de perspectivas inevitavelmente marcadas por sua própria cultura, seus próprios valores, seus próprios interesses. Além disso, muitos objetos têm significados espirituais, cerimoniais ou culturais específicos para as comunidades de origem, significados que são ignorados, desrespeitados ou mesmo violados quando esses objetos são tratados como meras peças de museu, disponíveis para contemplação estética de visitantes ocasionais, destacados de seus contextos rituais. Objetos sagrados para povos indígenas, por exemplo, não foram feitos para serem vistos por qualquer um, não foram feitos para serem expostos em vitrines; sua mera exibição pública pode ser uma profanação, uma violência simbólica. E o argumento do acesso global, tão caro aos defensores dos museus universais, também é problemático, porque o acesso físico aos grandes museus do Norte global é, na prática, restrito a uma minoria privilegiada da população mundial, aquela que pode arcar com os custos de viajar até Londres, Paris ou Nova York, que tem passaporte, que tem tempo, que tem dinheiro. A crítica de Linda Tuhiwai Smith (1999) articulou de forma pioneira a perspectiva descolonial sobre o problema. Sua tese é a de que a pesquisa e a custódia institucional de objetos e documentos indígenas são elas mesmas formas de colonialismo epistêmico quando realizadas sem o consentimento, a participação e o controle das comunidades envolvidas.

Uma posição mais equilibrada, me parece, e que tem sido adotada por algumas instituições e em alguns casos, é a que reconhece que não há uma res-

posta única que valha para todos os casos, que cada pedido de repatriação precisa ser examinado em seu contexto específico, levando em conta uma série de fatores, como, por exemplo, como o objeto foi adquirido originalmente, se foi saqueado, comprado, trocado, doado; qual o seu significado para a comunidade de origem hoje, se ainda tem relevância espiritual, cultural, política; se essa comunidade tem condições de preservá-lo adequadamente, se há infraestrutura, recursos, pessoal treinado; e se arranjos alternativos, como empréstimos de longo prazo, a propriedade compartilhada, a co-gestão, a digitalização e o acesso virtual, podem satisfazer os interesses e direitos de ambas as partes de forma criativa. O que não é aceitável, eticamente, é simplesmente ignorar os pedidos, tratá-los como uma ameaça à integridade das coleções ou como uma exigência descabida, sem entrar no mérito das questões de justiça histórica que eles levantam, sem dialogar com as comunidades, sem considerar a possibilidade de reparação.

O marco ético mais relevante para orientar essas negociações é o princípio do consentimento livre, prévio e informado (FPIC, *Free*, *Prior and Informed Consent*), consagrado na Declaração das Nações Unidas sobre os Direitos dos Povos Indígenas (2007), que estabelece que nenhuma decisão que afete bens culturais indígenas pode ser tomada sem a consulta e o consentimento das comunidades afetadas.

Outra questão ética fundamental, e que tem se tornado cada vez mais premente com os movimentos de descolonização dos arquivos e com a crescente vocalização de grupos historicamente silenciados, diz respeito à autoridade interpretativa sobre os documentos. Quem tem o direito de interpretar o passado a partir dos arquivos? Quem tem autoridade para decidir o que os documentos significam, que histórias eles contam, que vozes eles representam? Durante muito tempo, essa autoridade coube quase que exclusivamente aos historiadores profissionais e aos arquivistas, ambos vistos como especialistas com o treinamento adequado, o conhecimento técnico e a imparcialidade necessária para lidar com as fontes de forma objetiva. Mas as comunidades representadas nos documentos, especialmente aquelas que foram historicamente marginalizadas, oprimidas, silenciadas ou mal representadas, têm cada vez mais reivindicado o direito de participar da interpretação de seu próprio passado, de ter voz na ma-

neira como suas histórias são contadas, de contestar narrativas que foram impostas de fora.

A reivindicação de autoridade interpretativa pelas comunidades articula-se com o conceito, na teoria do arquivo, de pós-custodialismo, i.e., a ideia de que a responsabilidade sobre os documentos não pertence exclusivamente à instituição que os guarda, mas é partilhada com os criadores e com as comunidades afetadas pelos documentos. O pós-custodialismo não é apenas uma posição administrativa, mas uma posição ética sobre a distribuição do poder de narrar o passado.

Não se trata, é claro, de negar o valor do conhecimento especializado, da formação acadêmica, do rigor metodológico. Trata-se de reconhecer que as comunidades podem ter perspectivas únicas, insights valiosos, conhecimentos tradicionais que complementam e, às vezes, corrigem a visão dos especialistas externos. Trata-se de reconhecer que os documentos não falam por si mesmos, que a interpretação é sempre um ato situado, que diferentes leitores trazem diferentes perguntas, diferentes sensibilidades, diferentes experiências para o encontro com o texto. E trata-se de reconhecer que, quando a interpretação diz respeito à vida, à cultura, à história de um povo, esse povo tem o direito de participar da conversa.

A formulação de que os documentos "não falam por si mesmos" conecta-se com o programa hermenêutico de Gadamer (1960/2008), segundo o qual toda interpretação é mediada pelo horizonte do intérprete, e a objetividade não consiste em eliminar esse horizonte, mas em torná-lo explícito e submetê-lo à fusão de horizontes (*Horizontverschmelzung*) com o horizonte do texto. A inclusão das comunidades na interpretação dos documentos que lhes dizem respeito é, nessa perspectiva, uma condição de uma hermenêutica mais adequada, não uma concessão extracientífica.

A filósofa Miranda Fricker (2007) nos ajuda a entender o que está em jogo aqui de forma mais precisa. Se os arquivistas e historiadores ignoram sistematicamente as interpretações que as comunidades fazem de seus próprios documentos, se desconsideram suas perspectivas, se não lhes dão voz nos processos de descrição, classificação e curadoria do material que diz respeito a elas, estamos diante de uma forma de injustiça epistêmica. A solução aqui passaria por práticas de curadoria colaborativa, por arquivos participativos, por projetos

de história oral e comunitária, em que as comunidades são envolvidas como parceiras ativas, não como meras informantes ou consultoras de última hora para legitimar decisões já tomadas. A injustiça hermenêutica tem uma dimensão especificamente arquivística que Fricker (2007) não explorou diretamente; quando certas práticas, experiências ou perspectivas não são documentadas sistematicamente, porque os sistemas de registro foram criados por grupos dominantes, a ausência arquivística perpetua a exclusão hermenêutica. Não há palavras para descrever o que não foi registrado. O trabalho de arquivos comunitários e de história oral é, nesse sentido, um trabalho de reparação hermenêutica.

Há também as questões éticas, sempre delicadas, relacionadas ao acesso e à privacidade. Os arquivos estão cheios de informações sensíveis sobre pessoas, como dados de saúde, orientação sexual, filiação política, envolvimento em processos judiciais, segredos de família, falências, divórcios, escândalos. Por um lado, há o princípio da transparência e do direito à informação, fundamental em sociedades democráticas. Pesquisadores precisam ter acesso a esses materiais para que possam fazer seu trabalho, para que a história possa ser escrita com base em fontes primárias, para que abusos do passado possam ser investigados e revelados, para que a sociedade possa aprender com seus erros. Por outro lado, há o direito à privacidade das pessoas envolvidas, especialmente quando ainda estão vivas, e o direito de suas famílias de verem respeitada a memória de seus entes queridos, de não verem expostos publicamente aspectos íntimos de suas vidas que não dizem respeito a ninguém.

A tensão entre transparência e privacidade é estrutural e não pode ser resolvida por uma hierarquia fixa de valores; ela precisa ser gerenciada caso a caso, com recurso a critérios de proporcionalidade. O marco legal mais desenvolvido para essa gestão, no contexto europeu, é o *Regulamento Geral de Proteção de Dados* (RGPD), que estabelece exceções específicas para o tratamento de dados pessoais para fins de arquivo público, pesquisa científica e histórica, condicionadas a salvaguardas adequadas.

Não há uma fórmula mágica, uma regra simples, um algoritmo que resolva esse conflito de uma vez por todas. O que existem são princípios que podem orientar a reflexão em cada caso concreto, ponderando os valores em jogo. Conforme já mencionamos, o princípio da minimização de dados sugere que se

deve coletar e reter apenas a informação estritamente necessária para os fins legítimos do arquivo, evitando acumular dados pessoais sem justificativa clara. O princípio da transparência exige que as políticas de acesso sejam claras, publicamente conhecidas, e aplicadas de forma consistente, para que os usuários saibam o que esperar e possam planejar suas pesquisas. O princípio do consentimento informado recomenda que, sempre que possível, as pessoas sejam consultadas sobre o uso de informações que lhes digam respeito, e que tenham a oportunidade de autorizar ou recusar esse uso. E o princípio da proporcionalidade lembra que as restrições de acesso devem ser proporcionais aos danos que se pretende evitar, e não medidas genéricas, excessivas, que acabam por inviabilizar a pesquisa legítima.

Esses quatro princípios, minimização, transparência, consentimento e proporcionalidade, correspondem de perto aos princípios da bioética principialista (*Beauchamp* e *Childress*, 1979/2019), adaptados ao contexto arquivístico. Essa correspondência sugere que a ética arquivística pode beneficiar-se metodologicamente da bioética, que desenvolveu mecanismos sofisticados de adjudicação entre princípios em conflito aplicados a casos concretos, como comitês de ética, protocolos de revisão e padrões de justificação pública de decisões.

No caso de documentos históricos que contêm informações sensíveis sobre pessoas já falecidas, a questão se torna ainda mais complexa e delicada. Alguns argumentam que, com a morte, cessa o direito à privacidade, e que a história tem o direito de conhecer a verdade sobre o passado, por mais incômoda, vergonhosa ou dolorosa que ela seja. Outros defendem que a memória das pessoas merece respeito, que a dignidade não cessa com a morte, e que a divulgação de certas informações pode causar sofrimento desnecessário a descendentes, pode manchar a reputação de famílias inteiras, pode ferir a memória daqueles que já se foram e não podem mais se defender. Mais uma vez, não há resposta fácil, e cada caso precisa ser avaliado em suas circunstâncias específicas. O problema do estatuto moral dos mortos é um dos mais debatidos na ética contemporânea. Se a morte extingue os interesses que o indivíduo tinha em vida, então a privacidade *post mortem* não pode ser fundamentada no interesse do falecido, mas apenas nos interesses dos vivos afetados pela divulgação, familiares, comunidades, grupos cujas reputações são afetadas. Essa perspectiva desloca o debate da privacidade do morto para os danos que a

divulgação pode causar aos vivos, o que altera substancialmente o quadro de avaliação. O que se pode fazer é estabelecer procedimentos claros e transparentes para a avaliação desses casos, envolvendo diferentes perspectivas (arquivistas, historiadores, juristas, representantes da comunidade, familiares) e ponderando cuidadosamente os valores em jogo.

Por fim, há a questão da responsabilidade do arquivista para com o futuro, uma responsabilidade que é quase sobre-humana de tão pesada. As decisões de avaliação documental (*appraisal*), aquelas escolhas difíceis sobre o que guardar e o que eliminar, o que preservar para sempre e o que deixar ir, são decisões que moldam, de forma concreta e irreversível, o que as gerações futuras poderão saber sobre o nosso tempo. Cada documento eliminado é uma voz que se cala para sempre, uma perspectiva que se perde, uma história que não poderá ser contada, um ângulo que não poderá ser explorado (e, talvez, uma imoralidade na ética de Floridi, 1999). Os arquivistas carregam, portanto, um peso enorme, o de antecipar o que pode ser importante para pesquisadores que ainda não nasceram, com perguntas que ainda não foram formuladas, sobre temas que talvez nem imaginamos, usando métodos que ainda não foram inventados. É uma responsabilidade que exige humildade, reflexão, abertura ao diálogo, e a consciência de que se estará sempre errando em alguma medida, que sempre se perderá algo, que a história será sempre uma construção a partir de ruínas. O que se pode fazer, o mínimo que se pode fazer, é documentar as decisões, registrar os critérios usados, explicar os fundamentos das escolhas, para que o futuro pelo menos saiba o que foi perdido e por quê, para que possa, talvez, contestar essas decisões, criticá-las, aprender com elas.

A avaliação documental é, nesse sentido, um exercício de ética da responsabilidade no sentido de Hans Jonas (1979/2006); uma responsabilidade para com aqueles que ainda não existem, cujos interesses não podem ser representados nos processos decisórios presentes, mas que serão diretamente afetados pelas escolhas feitas hoje. A única resposta responsável a essa assimetria é a humildade epistêmica na avaliação, a consciência de que os critérios de relevância do presente não coincidirão com os do futuro, e a documentação transparente das decisões.

5. Considerações Finais

Ao longo deste capítulo, percorremos um caminho que nos levou às questões mais fundamentais sobre a natureza dos arquivos, sobre o conhecimento que podemos extrair deles e sobre as responsabilidades éticas que temos para com eles e para com todos os que são afetados por eles. Vimos que o documento de arquivo é uma entidade complexa, com uma ontologia dual que o situa entre o físico e o lógico, entre o material e o semântico. Sua identidade não reside apenas no seu suporte físico, que pode mudar com o tempo e com as intervenções de restauração, mas também, e talvez principalmente, nas relações contextuais que o ligam a outros documentos e ao seu fundo de origem. A autenticidade, tão central para a função probatória dos arquivos, revelou-se uma propriedade que depende não apenas de características intrínsecas do documento, mas de uma cadeia contínua de custódia e de práticas institucionais que atestam sua integridade ao longo do tempo, e de um trabalho de interpretação que envolve múltiplas camadas de evidência.

Na epistemologia, vimos que a confiança nos arquivos como fontes de conhecimento sobre o passado não é automática, nem pode ser dada como garantida. Ela depende de procedimentos rigorosos de verificação, de uma compreensão profunda dos contextos de produção e transmissão dos documentos, e de uma crítica constante das fontes. O princípio da proveniência e o da ordem original não são apenas regras de organização, meras convenções administrativas; são dispositivos epistemológicos que nos permitem ler os documentos à luz de suas relações, que nos ajudam a contextualizar e interpretar corretamente cada peça individual, que nos protegem contra os erros e distorções que viriam do isolamento dos documentos de seu contexto vital.

Na ética, nos confrontamos com dilemas que não admitem soluções fáceis, que exigem ponderação, diálogo, sensibilidade histórica e abertura ao outro. A tensão entre a preservação do patrimônio global e os direitos das comunidades de origem sobre seus bens culturais, entre a transparência democrática e a proteção da privacidade individual, entre o direito à informação e o respeito à memória dos que se foram, entre a autoridade dos especialistas e a participação das comunidades, todas essas são tensões que os arquivistas precisam navegar com sabedoria, com humildade, com a disposição de revisar as próprias

decisões à luz de novos argumentos e de novas vozes. Não há manuais que resolvam essas questões de antemão; há apenas princípios que orientam a reflexão e a coragem de decidir mesmo na incerteza.

O pluralismo prático que emerge da ética arquivística não é uma forma de relativismo, pois não afirma que todas as decisões são igualmente defensáveis, mas que a justificação de uma decisão é sempre situada, dependente de contextos, de relações de poder, de acordos institucionais e de valores que precisam ser tornados explícitos e criticáveis. A filosofia tem aqui a função prática de tornar visíveis os pressupostos que operam silenciosamente nas decisões arquivísticas cotidianas.

O pluralismo que emerge dessas reflexões, e que ecoa fortemente os capítulos anteriores deste livro, desde a antropologia até a ciência da informação, reconhece que a arquivologia não pode ser reduzida a uma única perspectiva, a um único método, a um único valor. Diferentes tradições arquivísticas, em diferentes países e épocas, privilegiaram diferentes aspectos do fenômeno arquivístico. Diferentes tipos de arquivo (públicos, privados, pessoais, comunitários, empresariais) exigem diferentes abordagens e diferentes ênfases. Diferentes contextos culturais e históricos demandam diferentes sensibilidades. A tarefa do arquivista reflexivo, do profissional que não apenas aplica técnicas mas também pensa sobre o que faz, não é aplicar fórmulas prontas, mas exercer um julgamento informado por uma compreensão filosófica clara dos conceitos em jogo, dos métodos disponíveis e dos valores em conflito.

Este capítulo, espero, tenha contribuído para essa compreensão, mostrando que a arquivologia, longe de ser uma disciplina menor, meramente técnica, está no centro de algumas das questões mais fascinantes e urgentes sobre memória, identidade, poder e justiça. Ele prepara o terreno de forma natural para os próximos capítulos. Na biblioteconomia, veremos como os livros e as coleções são organizados para acesso e disseminação, enfrentando questões análogas sobre identidade, classificação, acesso e preservação. Na museologia, veremos como objetos se tornam portadores de significado cultural, como são selecionados, exibidos e interpretados, acrescentando uma nova camada de complexidade à nossa compreensão da memória e do patrimônio. A jornada continua, dos arquivos às bibliotecas, das bibliotecas aos museus, sempre em busca de entender como lidamos com os vestígios do passado, como construí-

mos, a partir deles, o conhecimento sobre nós mesmos, e como negociamos as tensões entre lembrar e esquecer, entre preservar e destruir.

7

Filosofia da Biblioteconomia

1. Por que uma Filosofia da Biblioteconomia?

Quando um bibliotecário seleciona um livro para fazer parte do acervo, quando decide como classificar uma obra usando a Classificação Decimal de Dewey, quando ajuda um usuário a encontrar informação sobre um assunto que ele mesmo mal sabe explicar, é fácil pensar que ele está apenas executando tarefas técnicas, aplicando os procedimentos aprendidos na faculdade, seguindo manuais e protocolos estabelecidos. Mas, se a gente parar para observar com mais atenção, cada uma dessas atividades cotidianas está repleta de questões filosóficas que merecem ser examinadas com cuidado.

O que é, afinal, uma biblioteca? É apenas um prédio cheio de livros organizados em estantes? Ou é algo mais, uma instituição com uma missão social, um espaço de mediação entre o conhecimento registrado e as pessoas que dele necessitam? E o que é um livro? É um objeto físico, com páginas, capa, lombada, ou é uma entidade abstrata, uma obra que pode se manifestar em diferentes suportes ao longo do tempo? Como o conhecimento deve ser organizado para que as pessoas possam encontrá-lo? A ordem que damos aos livros nas estantes reflete uma estrutura objetiva do saber ou é apenas uma convenção útil, uma dentre muitas possíveis? E que responsabilidades éticas os bibliotecários carregam? Eles devem ser neutros em relação ao conteúdo que disponibilizam, ou têm o dever de promover certos valores em detrimento de outros? Como equilibrar a liberdade intelectual dos leitores com a responsabilidade social de não disseminar discursos de ódio ou desinformação?

As perguntas aqui formuladas correspondem à tripartição clássica da filosofia aplicada à informação, a saber, (1) questões ontológicas sobre o que é uma biblioteca e o que é um livro; (2) questões epistemológicas sobre como o co-

nhecimento deve ser organizado e recuperado; (3) questões éticas e normativas sobre quais responsabilidades os profissionais carregam. Essa estrutura revela que a prática bibliotecária pressupõe, implicitamente, posições filosóficas em cada um desses domínios, posições que a reflexão filosófica pode tornar explícitas, examinar criticamente e, se necessário, revisar.

Não há experimento científico, não há estatística de empréstimos, não há pesquisa de satisfação de usuários que responda a essas perguntas. Elas exigem outro tipo de investigação. Exigem que a gente pare para pensar sobre os conceitos fundamentais que estruturam a prática bibliotecária, sobre os pressupostos que estão silenciosamente orientando cada decisão, sobre os valores que estão em jogo em cada escolha. Exigem, em suma, filosofia.

A biblioteconomia, como campo do conhecimento e como prática profissional, tem uma história longa e rica. Desde as antigas bibliotecas da antiguidade, como a de Alexandria, passando pelas bibliotecas monásticas da Idade Média, onde monges copistas preservavam o saber clássico, até as bibliotecas públicas modernas, que se intencionam espaços democráticos de acesso à informação para todos, a questão sobre como organizar, preservar e disponibilizar o conhecimento sempre esteve presente. Mas foi apenas no século XIX e, principalmente, no século XX que a biblioteconomia se consolidou como uma disciplina com teorias, métodos e princípios próprios.

A história da biblioteca como instituição pode ser lida como uma história da materialização de diferentes epistemologias; por exemplo, a biblioteca alexandrina como projeto enciclopédico de totalização do saber; a biblioteca monástica como seleção teológica do que merece ser preservado; a biblioteca pública moderna como concretização de um ideal iluminista de democratização do conhecimento. Cada modelo institucional pressupõe uma teoria implícita sobre o que é o saber, quem tem direito a ele e como deve ser organizado.

Nomes como Melvil Dewey, que criou a Classificação Decimal de Dewey no final do século XIX, e S. R. Ranganathan (1931), o bibliotecário e matemático indiano que formulou as cinco leis da biblioteconomia na década de 1930, são figuras centrais nessa história. Dewey acreditava que era possível criar um sistema universal de classificação que organizasse todo o conhecimento humano de forma lógica e hierárquica. Ranganathan (1931), por sua vez, propôs uma abordagem mais flexível, baseada em facetas, e enunciou princípios que até

hoje orientam a reflexão sobre o papel social das bibliotecas, como "livros são para usar", "a cada leitor seu livro", "a cada livro seu leitor", "economize o tempo do leitor" e "a biblioteca é um organismo em crescimento", conforme já apresentamos anteriormente.

O que esses pioneiros fizeram, cada um à sua maneira, foi construir teorias sobre o que é uma biblioteca e como ela deve funcionar. Mas essas teorias, por mais influentes que tenham se tornado, não são verdades eternas e indiscutíveis. Elas incorporam pressupostos filosóficos que podem e devem ser examinados. A ideia de que é possível uma classificação universal do conhecimento, por exemplo, pressupõe uma certa visão sobre a natureza do saber humano. A ideia de que a biblioteca deve ser neutra em relação aos conteúdos que disponibiliza pressupõe uma certa posição sobre o papel social da instituição. E a ideia de que o bibliotecário é um mero aplicador de técnicas, um técnico neutro, pressupõe uma certa concepção sobre o que significa ser um profissional.

A crítica à pretensão de neutralidade técnica do bibliotecário conecta-se com a crítica mais ampla à figura do "*expert* neutro" nas ciências sociais aplicadas. Miranda Fricker (2007) argumenta que a mediação epistêmica é sempre valorativa, pois selecionar, organizar, classificar e disponibilizar conhecimento são atos que distribuem poder cognitivo e que, portanto, têm uma dimensão política inevitável.

Este capítulo é um convite para pensarmos juntos sobre todas essas questões. Vamos começar pela metafísica, tentando entender que tipo de entidade é uma biblioteca e que tipo de entidade é um livro. Depois, vamos para a epistemologia, para explorar como o conhecimento é organizado, classificado e tornado acessível, e qual o papel do bibliotecário nesse processo. E vamos terminar na ética, com as perguntas mais delicadas sobre neutralidade, censura, privacidade e o papel social das bibliotecas na democracia. Não se trata de encontrar respostas definitivas, mas de aprender a formular melhor as perguntas.

2. Metafísica da Biblioteconomia

A metafísica, como vimos ao longo deste livro, é o ramo da filosofia que investiga a natureza fundamental da realidade, i.e., que tipos de coisas existem, quais são suas propriedades essenciais, como se relacionam entre si. Quando

aplicamos essas perguntas à biblioteconomia, nos deparamos com questões que podem parecer estranhas à primeira vista, mas que são absolutamente centrais para a compreensão do que é uma biblioteca e do que é um livro.

O que é, afinal, uma biblioteca? A resposta pode parecer óbvia. Uma biblioteca é um lugar onde se guardam livros. Mas, se a gente pensar com mais cuidado, verá que essa definição simples não capta a complexidade do fenômeno. Uma biblioteca não é apenas um depósito de livros, como um armazém guarda mercadorias. Uma biblioteca tem uma missão, um propósito, uma função social. Os livros não estão ali apenas para serem guardados; estão ali para serem usados, consultados, emprestados, lidos. E a maneira como eles são organizados, classificados e disponibilizados não é neutra; reflete escolhas sobre o que é importante, sobre como o conhecimento deve ser acessado, sobre quem merece ter acesso a quê.

Podemos pensar a biblioteca em pelo menos três níveis ontológicos diferentes, cada um revelando aspectos distintos da sua natureza. No nível mais básico, a biblioteca é um conjunto físico de objetos. Livros, periódicos, manuscritos, mapas, partituras, CDs, DVDs, todos ocupando espaço físico em estantes, prateleiras, armários, salas. Essa é a biblioteca como depósito, como coleção de itens materiais. Mas, se pararmos aí, perderemos exatamente o que faz de uma biblioteca algo diferente de um armazém ou de uma loja. Uma coleção particular, por mais valiosa que seja, por mais rara que sejam seus livros, continua sendo uma expressão dos gostos e interesses do seu colecionador. Reflete sua subjetividade, suas idiossincrasias, suas limitações. Uma biblioteca institucional, pública ou universitária, por outro lado, aspira a algo diferente. Ela não é a expressão de uma mente individual, mas a materialização de uma função social. Sua identidade não é definida pela psicologia do seu fundador, mas pela sua missão declarada de educar, informar, preservar, promover a cultura, garantir o acesso democrático ao conhecimento.

Isso nos leva ao segundo nível ontológico, a saber, a biblioteca como instituição social. Nessa perspectiva, a biblioteca é um artefato cultural complexo, uma entidade que existe na interseção entre o físico e o ideal, entre o material e o normativo. Ela é, ao mesmo tempo, um edifício com estantes e mesas de leitura, e uma rede de regras, práticas, convenções e expectativas sociais. A biblioteca pública, em particular, é uma das grandes invenções democráticas da

modernidade. Ela encarna um compromisso com o acesso equitativo ao conhecimento, com a educação continuada, com a formação de cidadãos críticos e informados. Sua existência pressupõe certos valores (a crença na importância da leitura, no direito à informação, na necessidade de espaços públicos de convivência e aprendizado), que são materializados na sua arquitetura, nos seus horários de funcionamento, nas suas políticas de empréstimo, na sua programação cultural. A biblioteca, portanto, não é um objeto natural, que existe independentemente de nós. É um fato institucional, no sentido que o filósofo John Searle (1995) dá ao termo, i.e., uma realidade que existe porque acreditamos coletivamente que ela existe e agimos de acordo com essa crença. A declaração "este é um livro da biblioteca" não descreve uma propriedade física intrínseca do objeto; ela atribui a ele um status funcional dentro de um sistema de regras socialmente aceitas. Significa que aquele livro pode ser emprestado, que deve ser devolvido em certa data, que está disponível para consulta, que faz parte de um acervo organizado.

A ontologia dos fatos institucionais de Searle (1995) implica que a biblioteca existe enquanto a crença coletiva em seu valor e em suas regras se mantém. O financiamento público, a legislação de depósito legal e os acordos de direitos autorais para acesso bibliotecário são as expressões jurídicas dessa manutenção coletiva. Quando essas estruturas enfraquecem, a biblioteca não apenas perde recursos, perde gradualmente a consistência ontológica que a distingue de um simples depósito.

Há ainda um terceiro nível ontológico, mais sutil e talvez mais fascinante, a saber, a biblioteca como sistema cognitivo, como dispositivo intelectual, como máquina de pensar em escala coletiva. Nessa acepção, a biblioteca é mais do que a soma dos seus livros. Ela é uma tecnologia da mente, uma extensão da nossa capacidade de processar informação e de construir conhecimento. A sua organização (o sistema de classificação que determina onde cada livro fica na estante, o catálogo que permite recuperá-lo por autor, título ou assunto, a sinalização que orienta o usuário pelos corredores) constitui uma interface entre o conhecimento humano registrado e a mente do leitor. E essa interface não é neutra, não é um mero canal transparente. Ela estrutura o universo informacional de maneiras específicas, criando caminhos privilegiados de acesso, agrupamentos temáticos, hierarquias implícitas de importância e relevância. A deci-

são de colocar a seção de ficção ao lado da de literatura brasileira, ou de separar as ciências exatas das humanidades, ou de classificar a filosofia antes da religião, não é logicamente necessária. Ela reflete uma concepção particular sobre a arquitetura do saber, sobre como as áreas do conhecimento se relacionam, sobre o que é mais fundamental e o que é derivado. A biblioteca, assim compreendida, é um mapa do conhecimento humano. E, como todo mapa, ela simultaneamente revela e oculta, facilita certas navegações e dificulta outras, torna visíveis certos territórios e invisibiliza outros. Sua ontologia é a de um mediador ativo, um agente que formata a informação antes mesmo que ela seja demandada.

A caracterização da biblioteca como "tecnologia da mente" é compatível com a teoria da mente estendida (*extended mind*) de Andy Clark e David Chalmers (1998), segundo a qual, se as ferramentas externas podem constituir, e não apenas auxiliar, processos cognitivos, então a biblioteca não é um repositório que a mente acessa, mas um componente do sistema cognitivo distribuído que inclui o leitor, os instrumentos de busca e os próprios documentos. A organização da biblioteca co-constitui o conhecimento que o usuário pode construir.

Se a biblioteca é uma entidade complexa, com múltiplas camadas ontológicas, o livro, sua unidade básica, não fica atrás em complexidade. O livro é uma entidade metafisicamente intrigante, que exemplifica de maneira paradigmática o problema da relação entre o universal e o particular, entre a forma e a matéria, entre o tipo e o token. Conforme indicamos, o filósofo Charles Sanders Peirce (1931-1935) introduziu a distinção entre tipo (*type*) e token (*token*). O tipo é uma entidade abstrata, universal, que pode ser instanciada ou exemplificada em múltiplos tokens concretos e particulares. A palavra "livro" escrita duas vezes neste parágrafo é um mesmo tipo instanciado em dois tokens físicos distintos, i.e., as marcas de tinta ou os pixels específicos que formam a palavra em cada ocorrência. Aplicando essa distinção ao nosso caso, poderíamos dizer que a obra *Memórias Póstumas de Brás Cubas* é um tipo abstrato, uma estrutura narrativa e linguística ideal, um conjunto de palavras em uma determinada ordem. Cada exemplar físico do livro, seja a primeira edição de 1881, guardada em algum arquivo raro, a edição de bolso da Companhia das Letras que você comprou numa livraria, o arquivo PDF disponível gratuitamente no site do Domínio Público, o audiobook narrado por um ator famoso, é um token desse tipo.

Essa análise resolve alguns problemas imediatos, mas acaba criando outros, ainda mais profundos. Ela nos permite dizer, com algum sentido, que destruir um exemplar de *Memórias Póstumas de Brás Cubas* não é destruir a obra, desde que outros tokens persistam. Isso concede ao livro uma forma de imortalidade condicional, uma existência que transcende a materialidade perecível de qualquer uma de suas encarnações. No entanto, a identidade do tipo é ela mesma problemática. O que garante que duas sequências de signos sejam tokens do mesmo tipo? Uma tradução da obra para o inglês é um token do mesmo tipo que o original em português? Uma adaptação para o cinema, que conta a mesma história mas com imagens em vez de palavras? Uma versão resumida para estudantes, que mantém o enredo mas corta muitas passagens? Uma edição crítica, que reproduz o texto original mas acrescenta notas de rodapé explicativas? Os limites da identidade tipológica são vagos, difusos, e dependem, em grande medida, de convenções e de decisões interpretativas. Para um linguista, a sequência exata de palavras em uma língua específica pode ser essencial para a identidade da obra. Para um teórico da literatura, a estrutura narrativa, os temas centrais, os personagens principais podem ser suficientes para identificar a mesma obra através de traduções e adaptações.

O problema dos limites da identidade tipológica é tratado por Nelson Goodman (1976) sob os conceitos de “autografia” e “alografia”. Obras autográficas (como pinturas) têm identidade vinculada ao objeto físico singular; obras alográficas (como textos literários) têm identidade definida por notação, admitindo tokens múltiplos do mesmo tipo. Goodman (1976) argumenta que a questão “é esta a mesma obra?” não tem resposta independente do sistema notacional adotado, o que implica que a identidade da obra literária é parcialmente convencional e institucional.

Além disso, a análise *tipo-token* parece privilegiar a dimensão lógico-semântica do livro em detrimento da sua dimensão física e histórica. Para a maioria dos fins de leitura, para o leitor comum que quer ter contato com a história, os tokens são de fato intercambiáveis. Não faz diferença ler a edição de bolso ou o PDF, o exemplar novinho ou um usado cheio de anotações. O conteúdo é o mesmo. No entanto, para o bibliófilo, para o colecionador de livros raros, para o historiador do livro, para a biblioteca que guarda um acervo patrimonial, o token específico é absolutamente insubstituível. A primeira edição de *Memórias*

Póstumas de Brás Cubas não é apenas um veículo para o conteúdo da obra; é um artefato histórico, um testemunho material de um momento específico na história da cultura brasileira. Suas características físicas, como o tipo de papel, a fonte tipográfica usada pela gráfica, as marcas do processo de impressão, os erros de revisão que foram corrigidos em edições posteriores, os *ex-libris* de antigos donos colados na guarda, as anotações manuscritas nas margens, as marcas de uso e de desgaste, tudo isso carrega informações que nenhuma reimpressão moderna, nenhuma cópia digital, pode transmitir integralmente. Este *token* particular tem uma biografia, uma história de pertencimento, de leitura, de circulação, uma proveniência que é única e irrepetível. Tem uma aura, no sentido benjaminiano do termo, uma presença única no tempo e no espaço. Reduzi-lo a uma mera instância de um tipo, a mais um exemplar intercambiável, é perder de vista a sua singularidade histórica e a sua materialidade significante. A metafísica do livro, portanto, precisa acomodar essa dualidade, essa tensão irredutível entre o livro como veículo de um conteúdo intelectual potencialmente replicável e o livro como objeto físico único, inscrito na história, portador de marcas materiais que são elas mesmas fontes de informação e de significado.

A "aura" benjaminiana, presença única no tempo e no espaço decorrente da autenticidade da tradição, tem análogo direto no conceito arquivístico de originalidade e no princípio de proveniência; a identidade epistêmica de um documento não é apenas seu conteúdo textual, mas sua inserção em uma cadeia de custódia e em um contexto de produção materialmente irrepetível. A digitalização preserva o conteúdo semântico mas destrói a aura, o que constitui simultaneamente uma perda epistêmica e uma transformação ontológica do objeto.

A ascensão do livro digital, do e-book, parece, à primeira vista, realizar uma espécie de "fantasia" platônica, a saber, a separação completa da ideia em relação à matéria, a libertação do conteúdo intelectual do seu invólucro corpóreo, a possibilidade de a obra existir como pura informação, fluindo livremente, desencarnada, disponível em qualquer lugar e a qualquer momento, independente de qualquer suporte físico particular. O e-book é frequentemente descrito, nessa narrativa, como "desmaterializado", como se tivesse rompido os laços com a matéria que sempre prenderam o livro ao mundo físico. Essa narrativa, por mais atraente que seja para os entusiastas da tecnologia, é, no entanto, uma ilusão metafísica. O livro digital não é imaterial; ele simplesmente possui

uma materialidade diferente, mais complexa, mais dependente de uma infraestrutura técnica extensa e sofisticada.

A narrativa da desmaterialização do texto digital é criticada por N. Katherine Hayles (2002) sob o conceito de "materialidade da informação", i.e., toda informação é inscrita em algum substrato físico, e as propriedades desse substrato afetam o que pode ser inscrito, transmitido e interpretado. O livro digital não liberta o texto da matéria; transfere sua dependência material de um substrato perceptualmente direto (papel, tinta) para um substrato tecnicamente mediado (chips, servidores, protocolos de rede), deslocando e complexificando a materialidade, sem eliminá-la.

A existência de um e-book requer uma infraestrutura colossal, que inclui servidores onde os arquivos são armazenados, redes de transmissão de dados que os transportam, dispositivos de leitura (e-*readers*, *tablets*, *smartphones*) que os exibem, softwares de decodificação que interpretam os formatos, padrões de formatação que garantem a legibilidade. Um arquivo EPUB é tão material quanto um livro de papel, pois consiste em configurações eletromagnéticas em um chip de memória, em padrões de magnetização em um disco rígido, em sinais elétricos trafegando por cabos de fibra ótica. O que muda é a escala, a relação com a percepção humana direta, a opacidade dessa materialidade para o usuário comum. Clicamos em um ícone na tela e o texto aparece magicamente, sem que tenhamos qualquer consciência dos fluxos de elétrons, das operações lógicas, das transmissões de dados que tornaram aquela aparição possível. A materialidade do livro digital se esconde por trás de interfaces que simulam a imaterialidade, que nos dão a ilusão de que o texto simplesmente está ali, disponível, sem a corporificação do papel.

Essa opacidade no nosso conhecimento do funcionamento do livro digital, no entanto, não torna sua materialidade menos real ou menos consequencial. Pelo contrário, ela introduz novas formas de fragilidade e de controle que são completamente diferentes daquelas do livro impresso. A materialidade do papel é direta e autônoma; uma vez impresso, o livro pode ser lido à luz do sol, em qualquer lugar, sem necessidade de dispositivos intermediários ou de fontes de energia externas. Sua deterioração é lenta e, em certa medida, previsível, dado que o papel oxida com o tempo, a tinta pode desbotar se exposta à luz, o mofo pode atacar em ambientes úmidos, mas são processos que ocorrem em escalas

de tempo relativamente longas e que podem ser retardados com cuidados adequados. A materialidade digital, por outro lado, é profundamente dependente e intermediada. Um arquivo digital não é diretamente acessível aos sentidos humanos; ele requer um dispositivo de hardware e software específico para ser interpretado e exibido. A sua preservação enfrenta uma dupla ameaça, que os técnicos chamam de obsolescência tecnológica. De um lado, as mídias físicas onde os bits estão gravados (disquetes, CDs, DVDs, HDs) se degradam com o tempo e se tornam ilegíveis; de outro, os formatos de arquivo se tornam obsoletos à medida que novos softwares deixam de suportá-los, tornando impossível a leitura mesmo que os bits estejam intactos. É a chamada "deterioração lógica", um fenômeno sem paralelo no mundo analógico.

Mais insidiosamente, a materialidade digital permite novas formas de controle sobre o acesso e o uso do texto que eram impossíveis, ou extremamente difíceis, no mundo do papel. Os sistemas de DRM (*Digital Rights Management*) são camadas materiais de restrição, códigos embutidos no arquivo que limitam a cópia, o empréstimo, a impressão, ou até mesmo a leitura após uma data de expiração. Eles transformam o que era uma propriedade natural do objeto físico (a posse de um exemplar conferia, por si só, certos direitos de uso, como ler, emprestar, vender, dar) em uma permissão revogável, concedida por uma entidade externa (a editora, a plataforma) e sujeita a condições. Nesse sentido, o livro digital não é mais desmaterializado, mas sim hiper-materializado em uma rede de dependências técnicas e jurídicas que condicionam e limitam a sua existência e a sua acessibilidade. A metafísica do livro na era digital, portanto, não é a metafísica do espírito livre e desencarnado, mas a de um objeto fantasmagórico, cuja existência e disponibilidade estão permanentemente amarradas a uma infraestrutura técnica complexa e a relações de poder econômicas e legais que fogem ao controle do leitor. A biblioteca digital, consequentemente, não é simplesmente uma coleção sem paredes, uma biblioteca expandida para o espaço virtual. Ela é uma instituição que precisa negociar com essa nova ontologia do livro, lutando para preservar os princípios de acesso, preservação e privacidade em um ambiente onde o objeto custodiado é, por natureza, mais fluido, mais facilmente controlável e mais perecível do que o seu antecessor de papel.

O DRM enquanto "hiper-materialização" pode ser articulado com o conceito jurídico de *first sale doctrine*. No regime do livro impresso, a primeira venda transfere ao comprador direitos de uso amplos (ler, emprestar, revender, destruir); no regime do livro digital com DRM, a primeira venda é substituída por uma licença de uso revogável e condicionada. Essa transformação não é apenas econômica, é uma mudança ontológica na relação entre leitor e texto, que passa de propriedade a concessão, de posse a acesso mediado por terceiro.

3. Epistemologia da Biblioteconomia

Se a metafísica pergunta o que a biblioteca e o livro são, a epistemologia pergunta como nós, seres humanos, nos relacionamos com eles do ponto de vista do conhecimento. Como o conhecimento é organizado nas bibliotecas? Como os usuários encontram o que procuram em meio a milhares, às vezes milhões, de itens? Que tipo de autoridade epistêmica as bibliotecas possuem? E qual é, afinal, o papel do bibliotecário na mediação entre o leitor e a informação?

A classificação bibliográfica, longe de ser um mero exercício técnico de arrumação, é uma das expressões mais ambiciosas e concretas de epistemologia aplicada que a humanidade já produziu. Sistemas como a Classificação Decimal de Dewey, criada por Melvil Dewey em 1876, ou a Classificação da Biblioteca do Congresso, desenvolvida nos Estados Unidos no início do século XX, são tentativas monumentais de criar um mapa universal do conhecimento humano, uma espécie de *mathesis universalis* organizada em estantes. Cada decisão tomada na criação desses sistemas, a divisão de todo o saber em dez classes principais, a ordenação hierárquica das disciplinas, a atribuição de números ou letras a tópicos específicos, a definição de quais assuntos são próximos e quais são distantes, carrega consigo pressupostos epistemológicos profundos sobre a natureza, a estrutura e o valor relativo dos diferentes domínios do saber.

A Classificação Decimal de Dewey, para ficarmos no exemplo mais conhecido e influente, reflete com notável transparência a visão de mundo da sua época e do seu contexto de origem, que era a América do final do século XIX, protestante, progressista, profundamente influenciada pelos ideais positivistas de ordenação e progresso do conhecimento. A sua estrutura decimal, que divide tudo em dez categorias, cada uma subdividida em dez, e assim sucessivamen-

te, não é logicamente necessária. É uma escolha arbitrária mas poderosa, que impõe uma ordem decimal, baseada no número de dedos das mãos, ao que poderia ser organizado de inúmeras outras formas. A decisão de fazer a Filosofia (classe 100) preceder a Religião (classe 200) e ambas precederem as Ciências Sociais (classe 300) não é um reflexo direto da estrutura da realidade, mas uma afirmação sobre a prioridade lógica ou fundacional da reflexão filosófica, seguida pela esfera do sagrado e depois pelo estudo das instituições humanas. A posição das Linguagens (classe 400) antes das Ciências Naturais (classe 500), tal como de alguma forma fizemos neste livro, sugere uma visão humanística onde a ferramenta fundamental da comunicação, a linguagem, precede e torna possível a investigação do mundo natural. O lugar da Tecnologia (classe 600) imediatamente após as Ciências Naturais aplica uma narrativa linear, tipicamente moderna, do conhecimento puro (a ciência) ao conhecimento aplicado (a tecnologia). As Artes (classe 700) e a Literatura (classe 800) aparecem como domínios da expressão criativa, da subjetividade, separados das ciências e das ciências sociais. E a História (classe 900) surge como o grande repositório da experiência humana no tempo, o registro de tudo o que aconteceu.

Essa estrutura não é, não pode ser, neutra. Ela canoniza certas divisões disciplinares que são, em grande medida, históricas e contingentes, produto do desenvolvimento das universidades europeias e norte-americanas. Ela marginaliza saberes não-ocidentais ou não-acadêmicos, forçando-os a se encaixarem em categorias que lhes são estranhas. Um conhecimento tradicional indígena sobre plantas medicinais, por exemplo, pode ser dilacerado entre a Botânica, a Medicina, a Farmacologia e a Antropologia, perdendo assim a sua integridade sistêmica, a sua coerência interna, a sua lógica própria. A classificação, portanto, não apenas organiza um conhecimento que já existe em algum lugar, pronto para ser arrumado. Ela também constitui o conhecimento de uma certa maneira para os fins da biblioteca. Ela determina o que será agrupado com o quê, quais associações intelectuais serão facilitadas e quais serão obscurecidas, quais caminhos de pesquisa serão pavimentados e quais serão bloqueados.

A crítica ao imperialismo epistêmico dos sistemas de classificação ocidentais foi sistematizada por Hope Olson (2002) no conceito de “poder de nomear” (*power to name*). A ideia é que classificar é atribuir identidade epistêmica a um objeto, e essa atribuição distribui a visibilidade e a invisibilidade de um modo

não neutro. Saberes indígenas, tradições de conhecimento oral e epistemologias de países em desenvolvimentos são marginalizados porque os sistemas dominantes foram construídos a partir de categorias disciplinares europeias do século XIX.

A crítica a essa rigidez, a essa pretensão de universalidade e neutralidade, levou ao desenvolvimento de abordagens epistemologicamente mais sofisticadas e flexíveis. O exemplo mais notável, conforme abordamos anteriormente, é a classificação facetada de S. R. Ranganathan (1931), desenvolvida na Índia nas décadas de 1930 e 1940. Ranganathan (1931), que era matemático de formação antes de se tornar bibliotecário, percebeu que o conhecimento humano não é uma árvore com ramos fixos e bem delimitados, mas sim um universo multidimensional, onde um mesmo objeto pode ser abordado a partir de múltiplas perspectivas igualmente legítimas. Um mesmo livro, por exemplo, pode ser classificado de acordo com o seu assunto principal, com a sua forma literária, com a sua língua original, com a época em que foi escrito, com o público a que se destina. A classificação tradicional, baseada em hierarquias fixas, força o classificador a escolher uma dessas dimensões como a principal, relegando as outras ao esquecimento ou a índices auxiliares.

A inovação de Ranganathan (1931) foi propor um sistema baseado em facetas (*facets*), que são categorias fundamentais e universais que podem ser combinadas de maneiras flexíveis para descrever a complexidade de qualquer assunto. Ele identificou cinco facetas básicas, resumidas na fórmula PMEST, a saber, Personalidade (o que distingue um assunto de outros), Matéria (o material de que algo é feito), Energia (a ação, o processo, a operação), Espaço (a localização geográfica) e Tempo (o período histórico). Um livro sobre "a construção de pontes de aço na Inglaterra do século XIX" poderia ser classificado combinando as facetas Energia (construção), Matéria (aço), Espaço (Inglaterra), Tempo (século XIX), e Personalidade (pontes). A notação desenvolvida por Ranganathan (1931), que usa pontos e dois-pontos para separar as facetas, permite uma flexibilidade enorme, criando números de classificação que podem crescer e se adaptar à complexidade do conteúdo.

Essa inovação não foi meramente técnica; foi profundamente epistemológica. Ela abandonou a ideia de um mapa fixo, estático e universal do conhecimento em favor de um sistema de coordenadas que pode ser ajustado e combinado

para representar diferentes perspectivas sobre o mesmo objeto, diferentes recortes da realidade. A segunda lei de Ranganathan (1931), "a cada leitor seu livro", é a expressão ética desse insight epistemológico, pois para ele não há uma única ordem correta, absoluta, que valha para todos os leitores e para todos os propósitos. O que há são diferentes ordens, relevantes para diferentes necessidades, diferentes interesses, diferentes contextos de busca. A tarefa da classificação, portanto, deixa de ser a de revelar a estrutura objetiva e última do mundo, passando a ser a de construir interfaces eficazes e flexíveis entre coleções heterogêneas e usuários com projetos intelectuais diversos e mutáveis.

A abordagem facetada de Ranganathan (1931) antecipa, em décadas, a análise de domínio proposta por Birger Hjørland (1995) e Hanne Albrechtsen (1995), de que o conhecimento não deve ser organizado segundo uma estrutura universal, mas segundo as práticas e os propósitos das comunidades epistêmicas que o produzem e utilizam. Diferentes domínios, como biologia, medicina, direito, organizam o saber de formas distintas, e qualquer sistema de classificação que ignore essa pluralidade acaba impondo injustificadamente uma epistemologia externa às comunidades que serve.

As cinco leis da biblioteconomia formuladas por Ranganathan (1931) em 1931, em seu livro "As Cinco Leis da Ciência da Biblioteca" (*The Five Laws of Library Science*), são um verdadeiro tratado filosófico sobre a natureza, o propósito e a epistemologia das bibliotecas, embora escritas na forma de aforismos simples e memoráveis. Ainda que já tenhamos falado delas em capítulos anteriores, vamos examiná-las com cuidado neste capítulo, pela relevância significativa que tem para o nosso tema aqui e porque cada uma dessas leis carrega implicações profundas para toda a reflexão filosófica sobre a biblioteconomia.

A primeira lei, "livros são para usar", parece uma obviedade, uma constatação trivial. Mas, no contexto em que foi formulada, era uma verdadeira declaração de guerra contra uma concepção de biblioteca que ainda vigorava em muitos lugares, a biblioteca como museu, como depósito, como lugar onde os livros são preservados, guardados, protegidos, mas não necessariamente colocados à disposição dos leitores. Ranganathan (1931) estava afirmando, contra essa visão, que o valor epistêmico dos livros, a sua razão de ser, só se realiza plenamente no uso. Um livro que ninguém lê, que fica trancado em uma estante, é como uma ferramenta que nunca é usada; existe, mas não cumpre a sua fun-

ção. O conhecimento, para Ranganathan (1931), é dinâmico, não estático. Ele existe para ser acessado, consultado, estudado, debatido, transformado. Essa lei tem implicações práticas profundas para tudo o que uma biblioteca faz, desde a localização do prédio (deve ser de fácil acesso), passando pelos horários de funcionamento (devem ser convenientes para os usuários), até as políticas de empréstimo (devem facilitar, não dificultar, a saída dos livros da biblioteca).

A primeira lei de Ranganathan (1931) tem precedente filosófico na distinção aristotélica entre potência e ato. O livro não consultado existe em potência como portador de conhecimento; apenas o uso atualiza essa potência. Do ponto de vista da epistemologia pragmatista de Dewey (1938/1986), o conhecimento só existe como tal na medida em que orienta a ação e resolve problemas, o que reforça a tese de Ranganathan (1931) de que a biblioteca só realiza sua função epistêmica quando seus acervos são efetivamente mobilizados.

A segunda lei, "a cada leitor seu livro", e a terceira, "a cada livro seu leitor", são complementares e expressam o mesmo princípio fundamental, o da importância de conectar as pessoas certas com os livros certos. A segunda lei olha do leitor para o livro, observando que todo leitor, com suas necessidades específicas, seus interesses particulares, seu nível de conhecimento, seu contexto, merece encontrar o livro que lhe é adequado. Não há um livro que seja o melhor para todos; o que há são livros apropriados para leitores específicos em situações específicas. Isso implica que a biblioteca deve conhecer seus usuários, entender suas necessidades, antecipar suas demandas. Por sua vez, a terceira lei olha do livro para o leitor, indicando que todo livro, por mais especializado, por mais raro, por mais difícil que seja, merece encontrar o seu leitor, a pessoa que dele precisa. Isso implica que a biblioteca não pode se contentar em apenas adquirir e guardar os livros; precisa também promovê-los, divulgá-los, torná-los conhecidos, garantir que cheguem às mãos de quem pode se beneficiar deles. Essas duas leis assumem, juntas, uma forte contextualidade epistêmica, pois o conhecimento relevante não é uma propriedade absoluta dos livros, mas algo que emerge da relação entre o livro e o leitor, entre o conteúdo e a necessidade, entre a oferta e a demanda.

A quarta lei, "economize o tempo do leitor", é um princípio pragmático de enorme alcance. Ela reconhece que a busca por informação tem custos, que o tempo e o esforço do leitor são recursos valiosos que não devem ser desperdi-

çados. Os sistemas de informação (a classificação, o catálogo, a sinalização, o atendimento) devem ser projetados para minimizar o esforço necessário para localizar os recursos relevantes. Isso reflete uma epistemologia que valoriza não apenas a verdade abstrata, a correspondência com o mundo, mas também a eficiência, a acessibilidade, a usabilidade do conhecimento. Um sistema que torna o conhecimento tecnicamente disponível, mas que exige um esforço desproporcional para ser acessado, é, do ponto de vista prático, um sistema falho.

A quinta lei, "a biblioteca é um organismo em crescimento", é talvez a mais profunda e a mais visionária de todas. Ela reconhece que o conhecimento não é uma entidade estática, acabada, imutável. Novos campos do saber emergem constantemente, as fronteiras entre as disciplinas se deslocam, novas perguntas são formuladas, novas respostas são encontradas. A biblioteca, como instituição que serve ao conhecimento, precisa ser capaz de crescer e de se adaptar a essas mudanças. Não pode se apegar a estruturas fixas, a classificações imutáveis, a modos de organização que se tornaram obsoletos. Ela precisa ser um organismo vivo, em constante evolução, capaz de incorporar o novo sem destruir o velho, de se reorganizar em função de novas demandas, de se reinventar sem perder a sua identidade. Essa lei rejeita qualquer epistemologia estática, qualquer pretensão de ter alcançado uma forma definitiva e final de organização do saber, em favor de uma visão evolucionária, dinâmica e aberta do conhecimento.

A quinta lei de Ranganathan (1931) pode ser lida à luz da epistemologia evolucionária de Karl Popper (1963), dado que nela o conhecimento cresce por conjecturas e refutações, não por acumulação linear. Uma biblioteca que cristaliza suas categorias e recusa a reorganização está, epistemicamente, bloqueando o crescimento do conhecimento que pretende servir. A metáfora do organismo é também relevante. Os organismos que não se adaptam ao ambiente perecem, o que implica que a rigidez classificatória é, a longo prazo, autofágica.

Se as leis de Ranganathan (1931) expressam a filosofia da biblioteconomia em um nível alto e abstrato, então a prática cotidiana do bibliotecário, especialmente no serviço de referência, é onde essa filosofia se torna concreta, onde os princípios gerais são aplicados a casos particulares. O serviço de referência é o coração epistemológico da biblioteca, o lugar onde a mediação entre o conhecimento registrado e o usuário acontece de forma mais intensa e personalizada.

O ato de uma consulta de referência raramente começa com uma pergunta bem formulada, clara e precisa. Mais frequentemente, ele se inicia com uma necessidade vaga, difusa, mal compreendida pelo próprio usuário. Alguém chega e diz "preciso de material sobre a Revolução Francesa". O que essa pessoa realmente quer? Pode ser que esteja buscando entender as causas econômicas do processo revolucionário. Pode ser que esteja interessada nos personagens principais, em Robespierre, em Danton, em Marat. Pode ser que queira comparar a Revolução Francesa com outras revoluções, como a Russa ou a Americana. Pode ser que esteja procurando fontes primárias, documentos da época, para um trabalho de pesquisa. Pode ser, simplesmente, que precise de um resumo básico para uma prova do colégio. O primeiro e mais sutil julgamento epistêmico do bibliotecário é justamente o de interpretar a necessidade real, a pergunta por trás da pergunta, o que o usuário realmente quer saber, mesmo que ele mesmo não saiba expressar com clareza.

Essa interpretação requer muito mais do que o conhecimento do acervo, do catálogo, dos sistemas de classificação. Requer uma espécie de sabedoria prática, de *phronesis* profissional, que combina experiência acumulada em atendimentos anteriores, empatia para se colocar no lugar do usuário, sensibilidade para captar pistas não-verbais, e a capacidade de fazer as perguntas certas para ajudar o usuário a clarificar a sua própria demanda. É um conhecimento tácito, no sentido que Michael Polanyi (1966) deu ao termo, dado que sabemos mais do que podemos dizer. O bibliotecário experiente simplesmente sabe o que o usuário precisa, mesmo que não consiga explicar exatamente como chegou a essa conclusão. E esse saber é adquirido através da prática refletida, não apenas do estudo de manuais.

Uma vez interpretada a necessidade, o bibliotecário mobiliza o seu conhecimento explícito do sistema, procurando o catálogo, a organização das estantes, as bases de dados assinadas, as referências rápidas, as fontes especializadas. Mas a boa mediação não se encerra com a simples localização de fontes potenciais. O bibliotecário de referência que se limita a apontar para uma estante e dizer "os livros estão ali" está fazendo apenas metade do trabalho. A mediação de qualidade, que realmente faz diferença na vida do usuário, avança para o ensino, para a capacitação. O bibliotecário não entrega o peixe; ensina a pescar. Explica a lógica por trás do sistema de classificação, indicando, por exem-

plo, que os livros sobre a história econômica da França estão nesta seção porque seguem a subdivisão geográfica após o número de história geral. Desvenda a sintaxe do catálogo, mostrando, por exemplo, como podemos usar aspas para buscar uma frase exata, ou o operador E para combinar termos. E, o mais crucial na nossa era de desinformação e sobrecarga informacional, guia o usuário na avaliação crítica das fontes encontradas.

Esta última função, a de educar para o pensamento crítico, tornou-se o coração da epistemologia biblioteconômica contemporânea. Num cenário onde a informação é produzida e disseminada em escala industrial, onde qualquer um pode publicar qualquer coisa na internet, onde algoritmos de recomendação criam bolhas e câmaras de eco, onde a desinformação se espalha mais rápido que a verdade, a simples localização da informação já não é suficiente. O bibliotecário atua como um treinador, um *coach* do pensamento crítico, ajudando o usuário a aplicar perguntas epistemológicas fundamentais a cada fonte que encontra. Quem é o autor e quais são as suas credenciais? Ele é realmente um especialista no assunto ou apenas alguém com uma opinião? A editora ou o veículo de publicação é respeitável, tem histórico de qualidade? Qual é a data da informação e ela ainda é relevante para o que estou pesquisando? A informação é corroborada por outras fontes confiáveis, ou é um caso isolado? Qual é o propósito declarado (e o não declarado) do documento? É informar, persuadir, vender um produto, difamar um adversário, entreter? Este treinamento não visa a inocular uma desconfiança cética generalizada, uma paranoia epistêmica que duvida de tudo e de todos. O objetivo é cultivar uma confiança informada, baseada na capacidade de discernir gradientes de credibilidade, de entender a informação dentro dos seus contextos de produção e circulação, de reconhecer os próprios vieses e limitações.

A função do bibliotecário como educador epistêmico conecta-se com o conceito de "literacia informacional" (*information literacy*), definido pela *American Library Association* (ALA, 1989) como a capacidade de reconhecer quando a informação é necessária e de localizá-la, avaliá-la e utilizá-la eficazmente. Essa competência não é neutra, pois envolve a internalização de critérios de avaliação que são eles mesmos epistemicamente situados, o que exige do bibliotecário uma boa reflexão sobre os padrões que transmite.

Dessa forma, a mediação do bibliotecário evolui de uma epistemologia da autoridade delegada (o bibliotecário como guardião e distribuidor do conhecimento canônico, aquele que sabe o que é bom e entrega ao usuário) para uma epistemologia da autonomia facilitada, em que o bibliotecário é parceiro no desenvolvimento das competências necessárias para navegar e construir conhecimento de forma independente. É uma mudança de paradigma que alinha a biblioteconomia com as exigências epistêmicas de uma sociedade democrática e digitalmente complexa, onde o cidadão precisa ser não apenas consumidor, mas também produtor crítico de informação.

4. Ética da Biblioteconomia

A ética da biblioteconomia pergunta, de forma direta e incômoda, sobre quais são as responsabilidades morais dos bibliotecários. Como devem equilibrar valores que frequentemente entram em conflito, como a defesa da liberdade intelectual e a promoção do acesso equitativo, a proteção da privacidade dos usuários e a necessidade de transparência, a neutralidade política e o compromisso com a justiça social? Que princípios devem orientar a seleção de materiais, a definição de políticas de acesso, a mediação entre leitores e informações?

O princípio da neutralidade constitui um dos pilares éticos mais veneráveis e, simultaneamente, mais contestados da biblioteconomia moderna. Formalizado em documentos fundamentais como a *Library Bill of Rights* da *American Library Association*, ele prescreve que as bibliotecas devem fornecer materiais que representem todos os pontos de vista sobre os assuntos, e que os bibliotecários devem evitar, rigorosamente, que as suas convicções pessoais, políticas, religiosas ou filosóficas influenciem a seleção do acervo ou o atendimento ao público. Este compromisso emerge historicamente como uma reação vital contra as múltiplas formas de censura (religiosa, política, moral) que ao longo da história tentaram controlar o que as pessoas podiam ler. É também um fundamento essencial para a confiança pública na instituição, já que a biblioteca só pode ser um espaço verdadeiramente aberto à investigação intelectual livre se for percebida como um território neutro, onde todas as ideias podem circular sem favorecimento ou perseguição.

No entanto, uma análise filosófica mais rigorosa, livre dos dogmas profissionais, revela que a neutralidade absoluta é um ideal impossível de ser plenamente realizado, e que a sua aplicação literal, sem reflexão crítica, pode levar a consequências eticamente perversas. A impossibilidade deriva da própria natureza da seleção bibliográfica. Os recursos são finitos, os orçamentos são limitados, o espaço físico nas estantes também. A decisão de adquirir um determinado livro implica, necessariamente, a decisão de não adquirir inúmeros outros. Este ato de seleção, por mais que seja guiado por critérios profissionais como relevância para a comunidade, demanda dos usuários, qualidade literária ou acadêmica, e equilíbrio da coleção, é inerentemente um julgamento de valor. Não existe um algoritmo objetivo, uma fórmula matemática, que possa determinar, livre de qualquer perspectiva cultural, social ou ideológica, o que constitui "qualidade" ou "relevância". A própria escolha de quais critérios usar (privilegiar a aquisição de best-sellers para atender à demanda popular, ou investir em obras de nicho acadêmico para servir a uma minoria de pesquisadores) reflete uma valoração sobre a missão da biblioteca, sobre o seu papel na comunidade, sobre quem ela deve servir prioritariamente.

A impossibilidade da neutralidade na seleção bibliográfica é análoga ao problema da teoria de que os dados são carregados de teoria (*theory-ladenness of observation*) na filosofia da ciência. Assim como a observação científica é sempre mediada por teorias que determinam o que conta como dado relevante, a seleção bibliográfica é sempre mediada por concepções sobre o que conta como conhecimento de qualidade. Reconhecer isso não invalida a aspiração à imparcialidade, mas a desloca de uma neutralidade substancial impossível para uma imparcialidade procedimental exigente.

O problema se agrava, e se torna um verdadeiro dilema ético, quando a biblioteca se confronta com conteúdos que parecem testar os limites da tolerância em uma sociedade democrática. Deveria uma biblioteca pública, em nome da neutralidade, adquirir e disponibilizar panfletos que propaguem abertamente discurso de ódio racial, que incitem a violência contra minorias? Deveria adquirir livros que neguem categoricamente o consenso científico sobre as mudanças climáticas, ou que propaguem teorias conspiratórias perigosas sobre vacinas? Deveria manter em seu acervo obras que defendam a superioridade de uma raça sobre outra, ou que justifiquem a discriminação de gênero? A aplicação

estrita e acrítica do princípio da neutralidade sugeriria que sim, desde que esses materiais representem um "ponto de vista" existente na sociedade, desde que não sejam ilegais na forma mais estrita.

No entanto, uma parte significativa da reflexão ética contemporânea argumenta que essa postura confunde, de forma perigosa, a neutralidade com a indiferença moral e epistêmica. Não se trata de defender a censura prévia ou a imposição de uma verdade oficial. Trata-se, dessa forma, de reconhecer que a biblioteca não opera no vácuo, mas sim em um mundo real marcado por desigualdades, discriminações e lutas por reconhecimento. Disponibilizar materiais que negam a humanidade de grupos inteiros, ou que disseminam informações falsas com potencial de causar danos concretos (como a recusa de vacinação), não é um ato neutro. É um ato que pode, na prática, legitimar discursos de ódio, amplificar a desinformação e causar sofrimento real a pessoas reais.

Aqui, a filosofia ética oferece ferramentas conceituais para um raciocínio mais matizado e contextualizado. A distinção entre tolerância e promoção é crucial. Uma biblioteca pode decidir não adquirir determinado material não porque esteja censurando um ponto de vista, mas porque, após uma avaliação cuidadosa, julgou que ele não atende aos padrões mínimos de qualidade informacional, rigor argumentativo, decência civil ou relevância para a sua missão educativa e de promoção do bem comum. Esta decisão pode ser fundamentada, por exemplo, em uma ética de tipo consequencialista, ao dizer que a disseminação de desinformação científica nociva ou de discurso de ódio causa danos reais e mensuráveis à sociedade, danos que a biblioteca, como instituição voltada ao bem público, tem a responsabilidade de não amplificar. Alternativamente, uma abordagem baseada na ética das virtudes poderia argumentar que o bibliotecário virtuoso cultiva a prudência (*phronesis*) para discernir, em cada caso, quando a representação de um "outro lado" serve ao debate democrático legítimo e quando serve apenas para dar uma plataforma a falsidades perigosas ou a discursos de ódio.

O dilema entre neutralidade e responsabilidade epistêmica tem precedente na distinção milliana entre liberdade de expressão e incitação ao dano, pois Mill (1859) sustenta em *Sobre a Liberdade* que a única justificativa para restringir a liberdade de expressão é a prevenção de dano concreto a terceiros. Aplicado à seleção bibliográfica, o princípio milliano fornece um critério não arbitrário para

distinguir seleção legítima de censura; por exemplo, a decisão de não adquirir é defensável quando o material disponibilizado causaria dano mensurável a membros da comunidade, e não meramente quando seu conteúdo é desagradável ou controverso.

Portanto, a ética da neutralidade bibliotecária, corretamente compreendida, não deve ser interpretada como um mandamento cego e mecânico para a inclusão acrítica de todo e qualquer conteúdo, independentemente do seu teor e das suas consequências. Ela deve ser entendida como um princípio regulativo que exige transparência, justificação pública e humildade epistemológica por parte dos profissionais. O bibliotecário deve ser transparente sobre os critérios de seleção que emprega. Deve estar preparado para justificar publicamente as suas decisões, não com apelos vagos e inescrutáveis ao "julgamento profissional", mas com argumentos claros baseados na missão da biblioteca, nas necessidades da comunidade, e em valores compartilhados. E deve cultivar a humildade intelectual para reconhecer que os seus próprios julgamentos são falíveis e revisáveis, mantendo-se aberto ao diálogo, à crítica e à reconsideração. Neste modelo, a neutralidade deixa de ser uma abstinência de julgamento, uma fuga da responsabilidade de escolher, para se tornar uma imparcialidade procedimental, um compromisso com processos decisórios justos, reflexivos, participativos e responsáveis, mesmo quando, e especialmente quando, o conteúdo das decisões inevitavelmente reflete valores.

Se a seleção pergunta o que disponibilizar, a ética do acesso pergunta a quem disponibilizar, e em que condições. O compromisso com o acesso universal e equitativo é a espinha dorsal moral da biblioteconomia pública. Ele traduz a crença fundamental de que a informação e o conhecimento não são mercadorias como quaisquer outras, mas sim recursos fundamentais para a autonomia individual, para a participação cívica, para a mobilidade social, para o exercício pleno da cidadania. No entanto, a noção de acesso equitativo evoluiu, nas últimas décadas, de uma concepção meramente formal e negativa (a ideia de que as portas da biblioteca estão abertas a todos, sem distinção) para uma concepção substantiva e complexa de justiça informacional, que leva em conta as múltiplas barreiras que de fato impedem as pessoas de acessarem e usarem a informação.

A evolução de uma concepção formal para uma concepção substantiva de acesso corresponde à distinção rawlsiana entre igualdade formal de oportunidades e equidade. A igualdade formal garante que as mesmas regras se apliquem a todos; a equidade exige que as circunstâncias de partida sejam corrigidas de modo que todos possam efetivamente exercer os direitos formalmente garantidos. Uma biblioteca que se limita à igualdade formal, portas abertas a todos, pode ser simultaneamente justa e excludente.

O acesso formal, puramente legal, é insuficiente em sociedades marcadas por profundas desigualdades estruturais. Uma biblioteca pública localizada no centro de uma grande metrópole pode estar formalmente aberta a todos, mas permanecerá de fato inacessível para um jovem morador de uma periferia distante, que depende de transporte público precário e demorado, ou que precisa trabalhar em horário comercial para sustentar a família. Pode estar repleta de livros acadêmicos e especializados, mas permanecerá intimidante e incompreensível para um adulto em processo de alfabetização, que não domina o código letrado. Pode oferecer computadores com acesso à internet, mas não a mediação necessária, o acolhimento e a paciência para que um idoso aprenda a navegar no ambiente digital com segurança e eficácia. A ética do acesso, portanto, exige uma ação proativa, deliberada e contínua para identificar e remover barreiras que vão muito além da mera admissão física no edifício.

Esta ética desdobra-se em múltiplas dimensões interconectadas. A dimensão econômica luta contra qualquer barreira financeira ao acesso, defendendo a gratuidade integral dos serviços básicos de empréstimo e consulta, e fornecendo acesso gratuito à internet e a recursos digitais para aqueles que não podem pagar por eles em casa. A dimensão física e arquitetônica exige que as bibliotecas sejam plenamente acessíveis a pessoas com deficiências motoras, visuais ou auditivas, e que estejam localizadas em pontos estratégicos do território, alcançáveis por transporte público, inclusive através de bibliotecas móveis ou pontos de serviço em bairros mais afastados. A dimensão intelectual reconhece que a igualdade de oportunidade requer equidade de capacitação. Isto implica desenvolver programas robustos e continuados de alfabetização informacional, mediação de leitura, apoio ao estudo e letramento digital, com especial atenção e foco nos grupos historicamente marginalizados e com menos acesso à educação formal.

É aqui, precisamente, que a ética do acesso se funde com o conceito de justiça epistêmica de Miranda Fricker (2007). Aplicando esse conceito à biblioteconomia, podemos ver como uma biblioteca que só coleciona obras escritas por autores consagrados pela academia, em linguagem hermética e técnica, e que desconsidera ou marginaliza as formas de saber oral, comunitário ou popular, pode estar perpetuando uma injustiça hermenêutica contra grupos inteiros, cujas experiências e conhecimentos não são capturados por aqueles moldes estreitos. A ética do acesso, portanto, demanda uma curadoria inclusiva, diversa e representativa, que busque ativamente incorporar vozes silenciadas, perspectivas marginalizadas e formas não hegemônicas de conhecimento, não como um favor ou um gesto de benevolência, mas como um imperativo de justiça que busca corrigir distorções históricas no registro do saber coletivo.

A injustiça hermenêutica de Fricker (2007) tem uma expressão institucional específica nas bibliotecas, pois acervos que não representam as experiências de grupos marginalizados privam esses grupos não apenas de informação, mas das ferramentas conceituais para nomear e comunicar suas próprias realidades. A curadoria inclusiva não é, portanto, apenas uma questão de representatividade estatística, mas de justiça epistêmica no sentido estrito.

O maior dilema ético no âmbito do acesso, no entanto, talvez seja o conflito, tantas vezes trágico, entre o princípio do acesso irrestrito à informação e outros direitos fundamentais, como o direito à privacidade, tal como no caso dos arquivos. A tradição bibliotecária de defender a confidencialidade dos registros de empréstimo é justamente considerada sagrada, pois protege a liberdade intelectual em sua dimensão mais íntima. Essa privacidade é a condição de possibilidade para uma exploração intelectual verdadeiramente livre, para a formação de opiniões autônomas, para o exercício da cidadania em uma sociedade democrática.

No entanto, essa defesa intransigente da privacidade pode ser posta à prova por demandas que, em tese, são igualmente legítimas. Uma investigação criminal pode precisar de acesso aos registros de empréstimo de um suspeito para encontrar pistas ou provas. Agências de segurança podem argumentar que, em nome da segurança nacional, precisam monitorar padrões de leitura que possam indicar planos terroristas. Não há solução algorítmica, não há fórmula pronta, para resolver este conflito de valores. A ética profissional exige,

mais uma vez, um julgamento ponderado, que normalmente inclina a balança fortemente em favor da privacidade, dada a sua importância constitutiva para a missão fundamental da biblioteca. Mas este julgamento também deve reconhecer a possibilidade, rara e sempre cercada de cuidados extremos, de exceções devidamente justificadas, com respaldo legal claro e escrutínio público. Qualquer exceção deve ser transparente, proporcional e sujeita a revisão, para evitar que casos particulares se transformem em precedentes que corrompam a confiança, que é o alicerce da liberdade intelectual na biblioteca.

O propósito último da ética bibliotecária, o seu horizonte normativo, é garantir que a biblioteca cumpra a sua função como uma instituição voltada para o bem comum. Este bem comum não é a mera soma dos interesses individuais dos seus usuários, nem a satisfação das preferências de mercado. É um valor coletivo, distinto, que emerge da própria existência de uma instituição pública dedicada à preservação, organização e disseminação democrática do conhecimento. Na era contemporânea, este compromisso com o bem comum enfrenta o seu desafio mais formidável na ascensão das grandes plataformas digitais comerciais e na consequente mediação algorítmica da informação.

As grandes plataformas (Google, Amazon, Facebook, YouTube, TikTok) tornaram-se os principais intermediários informacionais para bilhões de pessoas em todo o mundo. Seus algoritmos de busca, recomendação e curadoria de conteúdo são projetados não para maximizar a verdade, a diversidade de perspectivas ou a profundidade do entendimento, mas sim para maximizar o engajamento do usuário e, consequentemente, a receita publicitária. Este modelo de negócio gera consequências epistêmicas perniciosas, amplamente documentadas por pesquisadores de diversas áreas, como a criação de câmaras de eco que reforçam preconceitos e isolam os usuários de visões divergentes, a amplificação desproporcional de conteúdos sensacionalistas, polarizantes e emocionalmente carregados, a difusão viral de desinformação e teorias da conspiração, e a consequente erosão de um espaço público compartilhado baseado em fatos e em discussões racionais.

Neste contexto desafiador, a biblioteca reafirma a sua razão de ser ética como um antídoto institucional à lógica predatória do mercado de atenção. O seu bem comum específico, a sua contribuição distintiva para a sociedade, é a criação e manutenção de um espaço informacional regido por valores não co-

merciais, por critérios de qualidade e relevância que não se reduzem ao número de cliques. Isto se manifesta de múltiplas maneiras concretas. Ao contrário do feed algorítmico opaco, cujos critérios são segredos comerciais, a curadoria intencional e transparente da biblioteca oferece coleções e serviços selecionados por profissionais com base em critérios públicos, justificáveis e orientados para as necessidades e interesses da comunidade, e não para o lucro de acionistas.

A organização física em estantes permite a navegação livre e a descoberta acidental de livros vizinhos, e a consulta a catálogos não personalizados, que não filtram os resultados com base no perfil do usuário, permitem descobertas acidentais, conexões improváveis e o contato com perspectivas inesperadas, algo que os algoritmos de personalização, ao nos darem sempre "mais do mesmo", sistematicamente eliminam. Assim, a biblioteca promove a serendipidade e o encontro com o diferente de um modo bastante especial. O valor epistêmico da serendipidade nas bibliotecas pode ser articulado com o conceito de "diversidade epistêmica" de Helen Longino (2002). Comunidades cognitivas saudáveis requerem o confronto com perspectivas alternativas, o que os algoritmos de personalização sistematicamente impedem, ao criar câmaras de eco. A organização física da biblioteca, onde o leitor que busca um livro encontra seus vizinhos de estante, é uma tecnologia de diversidade epistêmica cuja eficácia os ambientes digitais personalizados não replicam.

E o mais crucial, a biblioteca serve como um espaço comunitário para a deliberação democrática. Em um mundo de debates online fragmentados, anônimos e frequentemente hostis, a biblioteca oferece um fórum físico onde pessoas reais podem se encontrar para discutir livros, participar de palestras, debater questões locais e engajar-se em conversas civilizadas com seus concidadãos. Ela modela e facilita um tipo de interação pública baseada no respeito mútuo, na escuta atenta e na referência a fontes compartilhadas e verificáveis.

Para cumprir esta missão no século XXI, no entanto, a biblioteca deve ir além do seu papel tradicional de guardiã de livros impressos. Deve tornar-se também uma instituição de alfabetização algorítmica (*algorithmic literacy*), educando o público sobre como funcionam os sistemas que moldam a sua dieta informacional diária, quais são os seus vieses inerentes, e como se pode navegá-los de forma mais crítica e autônoma. Deve também desenvolver e apoiar

ativamente infraestruturas digitais alternativas e eticamente orientadas, como repositórios de acesso aberto, plataformas de publicação não comerciais, e ferramentas de busca que priorizem a privacidade e a relevância pública sobre o lucro privado.

Em última análise, a ética da biblioteconomia contemporânea é uma ética de resistência e de reafirmação. Resistência à colonização de todas as esferas da vida, inclusive a vida da mente, pela lógica implacável do mercado e da vigilância. Reafirmação dos valores republicanos da cidadania informada, do valor intrínseco da cultura e da sabedoria, e da necessidade ineliminável de instituições públicas dedicadas não à satisfação de preferências individuais efêmeras, mas à criação das condições para uma vida comum mais reflexiva, mais justa e mais livre. O bibliotecário, neste projeto, não é um mero administrador técnico de coleções, mas um guardião ativo e consciente deste bem comum informacional, um profissional cujo trabalho diário é, na sua essência, um ato de manutenção das bases epistemológicas e éticas de uma democracia que aspire a ser mais do que a soma dos seus cliques e das suas bolhas.

5. Considerações Finais

A investigação filosófica da biblioteconomia, ao longo deste capítulo, revelou que a disciplina está profundamente entrelaçada com questões fundamentais sobre a natureza dos livros e das bibliotecas, sobre a organização do conhecimento e sobre a responsabilidade ética dos profissionais da biblioteconomia. Longe de ser uma mera técnica de gestão de acervos, a biblioteconomia incorpora e enfrenta problemas filosóficos substantivos que resistem a soluções puramente procedimentais ou mecânicas.

Metafisicamente, vimos que a biblioteca é uma entidade complexa e multifacetada, que pode ser compreendida em pelo menos três níveis distintos mas inter-relacionados, como coleção física de objetos, como instituição social portadora de valores e como sistema cognitivo que estrutura o acesso ao conhecimento. O livro, a sua unidade básica, possui uma natureza dual, sendo ao mesmo tempo um tipo abstrato (a obra) e um *token* material (o exemplar físico). Essa dualidade gera tensões que se tornam ainda mais agudas na era digital, onde a promessa de desmaterialização do livro esbarra na realidade de novas e

complexas formas de materialidade, dependentes de infraestruturas técnicas e jurídicas que condicionam o acesso e a preservação.

Epistemologicamente, vimos que os sistemas de classificação, como a Classificação Decimal de Dewey, são muito mais do que meras ferramentas de arrumação. Eles são epistemologias aplicadas, que incorporam teorias específicas sobre a estrutura do conhecimento humano, sobre as relações entre as disciplinas, sobre o que é fundamental e o que é derivado. As cinco leis de Ranganathan (1931), por sua vez, expressam uma filosofia profunda e visionária sobre a natureza dinâmica, contextual e pragmática do conhecimento, antecipando debates que só se tornariam centrais com o advento do mundo digital. A mediação do bibliotecário, especialmente no serviço de referência, revelou-se um processo epistêmico complexo, que combina conhecimento tácito e explícito, e que evolui de uma epistemologia da autoridade delegada para uma epistemologia da autonomia facilitada, formando leitores críticos e autônomos.

Eticamente, vimos que a neutralidade absoluta é um ideal inatingível e, se interpretada de forma literal, pode levar a consequências moralmente problemáticas. A distinção entre seleção legítima e censura arbitrária, longe de ser óbvia, exige transparência, justificação e um constante exercício de julgamento ponderado. A ética do acesso deve ir além da igualdade formal para abraçar uma justiça informacional substantiva, que reconheça e combata ativamente as múltiplas barreiras (econômicas, físicas, intelectuais, epistêmicas) que impedem o acesso pleno à informação. Na era da vigilância digital e do capitalismo de plataforma, a biblioteca se reafirma como uma instituição do bem comum, um antídoto necessário à lógica do mercado de atenção, um espaço de resistência e de reafirmação dos valores democráticos fundamentais.

O pluralismo que emerge desta análise, e que ecoa fortemente os capítulos anteriores deste livro, desde a antropologia até a arquivologia, reconhece que a biblioteconomia não pode ser reduzida a uma única perspectiva teórica, a um único método de trabalho, a um único valor ético. Diferentes tipos de biblioteca (públicas, universitárias, escolares, especializadas, comunitárias) exigem diferentes ênfases e diferentes abordagens. Diferentes comunidades de usuários têm diferentes necessidades, diferentes contextos, diferentes valores. A tarefa do bibliotecário reflexivo, do profissional que não apenas executa tarefas mas pensa criticamente sobre o que faz, não é aplicar fórmulas prontas ou seguir

manuais cegamente. Mas, antes, exercer um julgamento informado, nutrido por uma compreensão filosófica clara dos conceitos em jogo, dos métodos disponíveis e dos valores em conflito, adaptando-se sempre às circunstâncias concretas de cada situação.

Este capítulo prepara o terreno, de forma natural, para o próximo, sobre a Filosofia da Museologia. Se as bibliotecas organizam livros para acesso e leitura, os museus expõem objetos para contemplação, aprendizado e experiência estética. Veremos como questões análogas, sobre identidade, interpretação, autenticidade e ética, se manifestam de formas específicas e fascinantes no contexto museológico. Veremos como objetos comuns se transformam em "semióforos", portadores de significado cultural, e como os museus negociam as tensões entre preservação e acesso, entre autoridade curatorial e participação comunitária, entre memória e esquecimento. O nosso caminho continua, do acesso à experiência, do livro ao objeto, do conhecimento a ser encontrado ao significado a ser contemplado.

8

Filosofia da Museologia

1. Por que uma Filosofia da Museologia?

Você já reparou como é estranha a experiência de entrar num museu? Não estranha no sentido de desconfortável, mas estranha no sentido de que algo muda na nossa relação com os objetos. Uma panela de barro que, na casa de alguém, seria apenas uma panela (útil, ordinária, talvez até meio velha), quando colocada numa vitrine, iluminada, com uma etiqueta ao lado dizendo "Vaso cerâmico, cultura Marajoara, século X", de repente se torna outra coisa. Ela ganha uma aura, um peso, uma importância que não tinha antes. A gente olha para ela de um jeito diferente. A gente não pensa em usá-la para cozinhar; a gente pensa na mão que a moldou, no tempo que passou, na cultura que a produziu. O que aconteceu aí? O que mudou, exatamente? O objeto físico continua sendo o mesmo, com a mesma argila, a mesma forma, as mesmas marcas. Mas algo se transformou. E é essa transformação, esse mistério, que está no coração da museologia.

Os museus são lugares estranhos também por outra razão. Eles estão cheios de objetos que não estavam destinados a estar ali. Aquela máscara africana não foi feita para ser admirada por turistas ocidentais numa sala climatizada; ela foi feita para ser usada em rituais, para assustar, para curar, para conectar os vivos com os mortos. Aquele crânio exposto numa vitrine não foi feito para ser visto; foi um dia parte de alguém, que talvez tenha sido enterrado com ritos e preces. Aquela estátua de um deus grego não foi feita para ser contemplada como obra de arte; foi feita para ser adorada num templo, para receber oferendas, para ser temida e amada. Os objetos de museu são, na sua maioria, objetos deslocados. Foram retirados dos seus contextos originais (contextos de uso, de culto, de vida) e recolocados num novo contexto, o contexto

museal, onde passam a ter novas funções, novos significados, novas vidas. Esse deslocamento não é inocente. Ele carrega histórias de violência, de pilhagem, de colonialismo. Mas carrega também histórias de preservação, de estudo, de admiração. Como lidar com essa ambiguidade? Compreender essa estrutura é o primeiro passo para uma filosofia da museologia.

Há ainda a questão do conhecimento. O que a gente aprende num museu? A gente aprende sobre o passado, sobre outras culturas, sobre a arte, sobre a ciência. Mas como aprendemos? Será que os objetos falam por si mesmos, ou são os curadores que falam por eles? Será que a disposição dos objetos numa sala, as cores das paredes, a iluminação, os textos explicativos, tudo isso molda o que a gente vê e o que a gente pensa? E se sim, será que existe um jeito "certo" de fazer isso? Ou toda curadoria é necessariamente uma interpretação, uma escolha entre muitas possíveis?

Os museus também são lugares de disputa. Disputa sobre quem tem direito aos objetos, sobre quem pode interpretá-los, sobre quem pode vê-los. Países que foram colonizados exigem de volta os seus tesouros espalhados pelos museus da Europa. Comunidades indígenas reivindicam os restos dos seus ancestrais, guardados em gavetas de museus de história natural para fins de pesquisa. Artistas contestam a forma como as suas obras são expostas. Visitantes questionam as narrativas que lhes são apresentadas. Essas disputas não são meramente políticas ou jurídicas; elas têm uma dimensão filosófica profunda, pois implicam desacordos sobre o que os objetos são, a quem pertencem por direito e que tipo de conhecimento produzem. A filosofia da museologia é o espaço em que essas dimensões se tornam explícitas e são submetidas a uma análise rigorosa.

Tudo isso são questões filosóficas. Não há experimento que resolva a disputa entre a Grécia e o Museu Britânico sobre os Mármores de Elgin. Não há técnica de conservação que nos diga se um objeto restaurado continua sendo o mesmo. Não há pesquisa de público que nos diga qual a interpretação correta de uma obra de arte. São questões que exigem que a gente pense, que a gente argumente, que a gente pese valores, que a gente decida. Este capítulo é um convite para pensar junto. Vamos entrar nesse mundo estranho e fascinante dos museus e tentar entender o que está em jogo. Vamos começar pela metafísica, pela pergunta sobre o que acontece com os objetos quando entram no museu.

Depois vamos para a epistemologia, para a pergunta sobre o que e como aprendemos ali. E vamos terminar na ética, com as perguntas mais espinhosas sobre propriedade, repatriação, justiça e memória.

A articulação dessas três dimensões não é apenas expositiva, elas se implicam mutuamente. A resposta que se dá à questão metafísica, o que é um objeto museal, determina o que contará como conhecimento legítimo sobre ele, o que por sua vez delimita as responsabilidades éticas de quem o custodia, exibe e interpreta.

2. Metafísica da Museologia

Vamos começar com uma cena. Imagine uma máscara africana, esculpida em madeira, com traços estilizados, olhos grandes, boca aberta. Durante séculos, ela foi usada em rituais noturnos numa aldeia. O dançarino a vestia e, durante a dança, ele não era mais ele mesmo; era o espírito que a máscara representava. A máscara tinha poder. Era temida e reverenciada. As crianças sabiam que não deviam tocá-la. Os velhos sabiam as histórias que ela carregava.

Agora imagine essa mesma máscara, cem anos depois, numa vitrine do Museu do Quai Branly, em Paris. Iluminada suavemente, com uma etiqueta discreta ao lado, onde se lê "Máscara, cultura Fang, Gabão, século XIX. Doada pelo missionário ß em 1923". Os visitantes passam, olham, admiram a forma, a textura, a cor. Alguém tira uma foto. Uma criança pergunta à mãe por que ele tem essa cara esquisita. A mãe lê a etiqueta e diz que "é uma máscara de um povo da África, querida. Eles usavam em rituais". O que aconteceu com a máscara? Ela é a mesma? Fisicamente, sim. A mesma madeira, as mesmas marcas de cinzel, a mesma pintura desbotada pelo tempo. Mas algo mudou, e mudou radicalmente. Na aldeia, ela era um agente, participava de rituais, tinha poder. Em Paris, é um objeto de contemplação, testemunho de uma cultura distante, "peça de museu". Perdeu o poder, ganhou significado. Deixou de ser coisa para agir e passou a ser coisa para significar.

O historiador polonês Pomian (1984) tem um nome para essa transformação. Ele chama esses objetos de *semióforos*, do grego *semeion* (signo) e *phoros* (que carrega). Os semióforos são objetos que não têm utilidade prática, mas que são dotados de significado. Eles representam algo que não está ali, algo

invisível, como o passado, o sagrado, o outro, o distante. Uma máscara ritual, um fóssil de dinossauro, uma pintura renascentista, uma carta de amor antiga, todos eles, no museu, são semióforos. Não servem para usar, servem para significar. Pomian (1984) desenvolve essa ideia no seu ensaio *Coleção*, mostrando como os objetos, ao entrarem em coleções, deixam de circular na economia do útil e passam a circular na economia do significado.

A noção de semióforo de Pomian (1984) articula-se com a distinção semiótica entre valor de uso e valor de signo, dado que, ao ser musealizado, o objeto não apenas perde sua função prática, mas adquire uma função representacional que passa a ser o fundamento de seu valor. Esse deslocamento não é meramente econômico; é ontológico, pois implica uma mudança no modo de ser do objeto, de coisa que age para coisa que significa.

A ideia de Pomian (1984) é poderosa porque nos ajuda a ver que a musealização não é apenas uma mudança de lugar, mas uma mudança de natureza. O objeto deixa de ser uma coisa no mundo para se tornar uma coisa sobre o mundo. Ele se torna representação, testemunho, documento. É por isso que os museus têm departamentos de conservação, políticas de aquisição, critérios de descarte. O que está sendo cuidado não é apenas a madeira, a tinta, o metal; é a capacidade do objeto de continuar significando.

Mas será que a transformação é tão radical assim? O historiador Stephen Bann (1984) desconfia. Para ele, a máscara não muda de natureza; o que muda é o contexto. Bann (1984), em *The Clothing of Clio*, argumenta que na aldeia, a máscara estava inserida numa rede de relações rituais, sociais e espirituais. Em Paris, está inserida numa rede diferente, de relações curatoriais, discursivas e estéticas. A máscara é a mesma, mas as suas relações são outras. É como uma pessoa que emigra, que continua sendo a mesma pessoa, mas a sua vida, as suas funções, os seus significados mudam porque mudou o contexto em que vive.

A posição de Bann (1984) aproxima-se do externalismo semântico, pois o significado de um objeto não é uma propriedade intrínseca, mas é determinado pelas relações que ele mantém com o ambiente, social, material, discursivo. Isso tem uma consequência importante para a museologia, a saber, a de que se o significado é relacional, então a musealização não destrói o significado original, mas substitui uma rede de relações por outra. A pergunta relevante torna-se

saber que tipos de relações o museu é capaz de reconstituir, e que tipos inevitavelmente perde.

Quem tem razão? Talvez os dois, talvez nenhum dos dois, talvez seja mais complicado. Susan Pearce (1992), museóloga britânica, propõe um caminho intermediário. Pearce (1992), em *Museums, Objects and Collections*, fala numa ontologia dupla dos objetos museais. Eles são, ao mesmo tempo, coisas físicas (com peso, textura, composição química, idade) e coisas culturais (com significado, valor, história, proveniência). Essas duas dimensões não estão separadas; elas se interpenetram. A materialidade do objeto não é um mero suporte para o significado; ela é parte do significado. A pátina do tempo, as marcas de uso, os arranhões, as restaurações, tudo isso conta uma história, tudo isso é significado incorporado.

A ontologia dupla de Pearce (1992) ressoa com o que a filosofia da arte chamou de propriedades estéticas emergentes, i.e., propriedades que dependem das propriedades físicas do objeto, mas que não se reduzem a elas. A pátina de uma escultura antiga não é apenas uma propriedade química (oxidação de determinados metais); é uma propriedade estética e histórica que informa sobre o envelhecimento do objeto, as condições de sua conservação, o tempo que o separa do presente. Essa emergência é o que torna a materialidade do objeto museal filosoficamente interessante.

Pense numa escultura grega que perdeu o nariz, os braços, que está desgastada pelo tempo. Esse desgaste não é apenas dano; é testemunho da passagem do tempo, da história da escultura, das mãos que a tocaram, dos séculos que passaram. Se um restaurador decidir reconstituir o nariz e os braços como eram originalmente, algo se perde; perde-se a história, ganha-se a forma original. É um *trade-off*.

Essa dualidade fica ainda mais complicada quando pensamos na restauração de um modo geral. Os museus estão cheios de objetos que foram mais ou menos restaurados. O esqueleto de dinossauro que você vê pode ter apenas 20% de ossos originais; o resto é resina, gesso, moldes. A pintura renascentista pode ter perdido grande parte da camada original de tinta em restaurações passadas. O vaso grego pode ser uma reconstituição a partir de cacos originais unidos por massa de preenchimento moderna. Esses objetos são os mesmos? Se são, em que sentido?

A teoria da restauração desenvolvida por Brandi (1963/2004) nos diz que o restaurador deve intervir de modo a ser reconhecível (a restauração não deve enganar) e reversível (deve ser possível desfazê-la no futuro). Esses dois princípios, reconhecibilidade e reversibilidade, podem ser lidos como tentativas de preservar a distinção ontológica entre o original e o acréscimo, mesmo quando ambos coexistem no mesmo objeto. A identidade do objeto é assim compreendida como estratificada, pois há o estrato original e os estratos das intervenções sucessivas, cada um com seu próprio estatuto (Brandi, 1963/2004, p. 47-63).

Esse é o velho paradoxo do navio de Teseu, que os filósofos discutem há milênios. O navio de Teseu era preservado como uma relíquia em Atenas. Com o tempo, as pranchas foram apodrecendo e sendo substituídas, até que nenhuma prancha original restava. O navio continuava sendo o mesmo? Se sim, o que garante a identidade? Se não, em que momento deixou de ser? E, ainda, se remontássemos o navio com as tábuas velhas, qual deles seria o original? No museu, a resposta não é teórica; ela tem consequências práticas. Um osso original tem um valor, uma informação, uma autenticidade que um molde não tem. Mas o molde permite que o público veja o animal completo, entenda a sua forma, se maravilhe com a sua grandeza. O original verdadeiro, se exposto, poderia se deteriorar rapidamente. Não há solução mágica, há decisões a tomar, e cada decisão envolve pesar valores diferentes.

A filósofa Lynne Rudder Baker (2007), em *The Metaphysics of Everyday Objects*, desenvolveu uma teoria que pode ajudar a pensar essas questões. Ela fala que um objeto físico (um punhado de argila) pode *constituir* um objeto de outro tipo (uma escultura) sem ser idêntico a ele. A escultura é feita daquela argila, mas não é a mesma coisa que a argila, pois a argila poderia existir sem ser escultura, e a escultura poderia, em princípio, ser feita de outro material. Dessa forma, o osso original constitui o fóssil, mas o fóssil pode, em certas circunstâncias, ser constituído também por moldes, desde que haja uma continuidade causal e a documentação adequada. A identidade não está na matéria, mas na história.

A teoria da constituição de Baker (2007) tem uma vantagem importante sobre as teorias da identidade por continuidade material, a saber, ela permite que objetos de tipos diferentes (um punhado de ossos e um esqueleto montado) sejam distintos sem que um deles seja "falso". O molde não é um osso original,

mas pode constituir genuinamente um fóssil para fins expositivos, desde que a distinção entre os dois estratos seja documentada e comunicada ao público. Isso tem implicações diretas para a ética da curadoria, dado que a transparência sobre o que é original e o que é reconstituição seria uma exigência filosófica que decorre da ontologia dos objetos museais.

Há ainda a questão da coleção. Uma coleção não é apenas um amontoado de objetos; é uma totalidade com uma lógica própria. Os objetos se relacionam entre si, se iluminam mutuamente, contam uma história em conjunto. Quando se retira um objeto de uma coleção, algo se perde que não se limita àquele objeto. A coleção fica mutilada, incompleta, como uma frase com uma palavra faltando. É por isso que os museus relutam tanto em vender ou doar peças das suas coleções. Não é só apego; é que a coleção tem uma integridade que precisa ser respeitada.

O estatuto ontológico da coleção é distinto do estatuto de cada peça que a compõe; ela é uma entidade emergente, cujas propriedades, o argumento que ela articula, a história que ela conta, as relações que ela evidencia, não se reduzem às propriedades dos objetos individuais. Isso torna a coleção um objeto filosófico interessante. Ela é construída por agentes humanos, mas, uma vez constituída, adquire uma integridade que transcende as intenções do construtor original.

Mas essa integridade é objetiva ou é construída? Para quem defende a visão construtivista, a coleção é uma invenção do curador, uma narrativa que ele impõe sobre objetos que, no fundo, não têm relação necessária entre si. O curador escolhe o que juntar, como juntar, que história contar. Outra pessoa, com outros interesses, poderia juntar os mesmos objetos de outra forma e contar uma história completamente diferente. A vantagem dessa perspectiva é que ela nos alerta para o poder dos curadores, para o fato de que as coleções não são neutras, que refletem escolhas, valores, visões de mundo. A desvantagem é que ela pode levar a um relativismo excessivo, como se qualquer narrativa fosse tão válida quanto qualquer outra, como se não houvesse fatos a respeitar.

Para quem defende a visão organicista, a coleção reflete uma história comum, uma proveniência partilhada. Os objetos que vieram de uma mesma escavação, de uma mesma coleção particular, de um mesmo contexto cultural, têm relações reais entre si, relações que existem independentemente do cura-

dor. O princípio da proveniência, tão caro à arquivologia, assenta nessa ideia, de que os objetos devem ser mantidos juntos porque as suas relações são parte do que eles são. A vantagem dessa perspectiva é que ela ancora a coleção na realidade, em fatos históricos, em relações que podem ser investigadas e documentadas. A desvantagem é que ela pode levar a um essencialismo ingênuo, como se as relações entre objetos fossem fixas e imutáveis, como se os objetos não pudessem ganhar novas relações, novos significados, ao longo do tempo.

Uma terceira via, que poderíamos chamar de construtivismo moderado, reconhece que as relações entre objetos são parcialmente dadas (pela história, pela proveniência, pelo contexto original) e parcialmente construídas (pela curadoria, pela interpretação, pelo arranjo expositivo). A tarefa do museólogo seria ser transparente sobre quais relações são de um tipo e quais são de outro, distinguindo o que a história impõe do que a curadoria propõe.

No fundo, o que a metafísica da museologia nos mostra é que os objetos de museu são entidades estranhas, híbridas, que vivem entre o material e o simbólico, entre o passado e o presente, entre o original e a cópia, entre o individual e o coletivo. Não há teoria que capture toda essa complexidade de uma vez. O que há são ferramentas conceituais que nos ajudam a pensar cada caso, a pesar cada decisão. E talvez seja isso que importa, a saber, a clareza sobre o que está em jogo.

3. Epistemologia da Museologia

Agora vamos mudar de perspectiva. Em vez de perguntar o que os objetos são, vamos perguntar o que a gente aprende com eles. Os museus se apresentam como lugares de conhecimento. As crianças vão em excursões escolares para “aprender sobre os dinossauros” ou “sobre o Egito antigo”. Os adultos visitam exposições para “conhecer a obra de um artista” ou “entender um período histórico”. Mas como funciona esse aprendizado? O que está acontecendo quando a gente aprende num museu?

A questão epistemológica central da museologia seria saber qual é a relação entre os objetos e o conhecimento que eles supostamente veiculam. Há duas posições extremas, o realismo ingênuo, que atribui ao objeto a capacidade de comunicar diretamente seu significado, e o construtivismo radical, que atribui

todo o significado ao discurso curatorial, e uma série de posições intermediárias que tentam articular a contribuição do objeto e a contribuição do intérprete.

Uma visão ingênua diria que os objetos falam por si mesmos. O fóssil está ali, e a gente olha e aprende que existiu um bicho enorme há milhões de anos. A pintura está ali, e a gente olha e aprende como Van Gogh via o mundo. Mas será que é tão simples assim? Pense no fóssil. Sozinho, ele não diz a sua idade, não diz o que comia, não diz como se movia, não diz por que foi extinto. Tudo isso a gente só sabe porque paleontólogos estudaram, compararam, dataram, interpretaram. O conhecimento não está no objeto; está na relação entre o objeto e uma rede de teorias, métodos, comparações, inferências. O objeto é mudo; quem fala por ele são os especialistas.

Pense na pintura de Van Gogh. Sozinha, ela não diz quem pintou, não diz quando, não diz por que, não diz o que significa. Se a gente vê um céu estrelado e rodopiante, a gente pode achar bonito, mas não sabe que é de Van Gogh, não sabe que ele pintou num hospício, não sabe que ele estava louco, não sabe que ele se matou pouco depois. Tudo isso é informação que vem de fora, que é acrescentada ao objeto pela curadoria, pelos textos de parede, pelos guias, pelos catálogos.

A dependência do conhecimento museal relativamente a contextos interpretativos exteriores ao objeto é uma necessidade estrutural. Charles Sanders Peirce (1931-1935) mostrou que todo signo só significa em relação a um interpretante, um sistema de regras, hábitos e convenções que determinam como o signo deve ser lido. No museu, o objeto é um signo cujo interpretante é fornecido pela curadoria; sem o interpretante curatorial, o objeto permanece opaco, um mero estímulo perceptivo sem significado determinado (Peirce, 1931-1935, vol. 2, p. 228-234).

A museóloga Eilean Hooper-Greenhill (1992) tem uma tese forte sobre isso. Em *Museums and the Shaping of Knowledge* (1992), ela argumenta que os museus não transmitem conhecimento; eles produzem conhecimento. As escolhas curatoriais, o que selecionar, como classificar, como dispor, o que destacar, o que omitir, que texto escrever, tudo isso molda o que o visitante vai aprender. Não há uma exposição neutra, que apenas mostre os objetos como eles são. Toda exposição é uma interpretação, uma narrativa, uma construção.

Hooper-Greenhill (1992) apoia-se explicitamente em Michel Foucault (1963/2003) para mostrar que os museus são dispositivos de produção de saber. Assim como a clínica médica constitui o objeto "doença" através de práticas de observação, classificação e registro, o museu constitui seus objetos, "arte", "artefato", "espécime", através de práticas curatoriais específicas. Mudar as práticas curatoriais não é apenas mudar o modo de apresentar objetos preexistentes; é constituir objetos diferentes (Hooper-Greenhill, 1992, p. 6-21).

Isso pode soar relativista, como se qualquer narrativa fosse tão válida quanto qualquer outra. Mas não é disso que se trata. Há fatos. Van Gogh existiu, pintou num hospício, morreu em 1890. Mas o significado desses fatos, a história que se conta a partir deles, a forma como se conectam com outras coisas, isso é matéria de interpretação. E diferentes interpretações são possíveis, e algumas são melhores do que outras, mais fundamentadas, mais coerentes, mais ricas, mais sensíveis aos detalhes.

O antropólogo Bruno Latour (1987), estudando como os cientistas produzem conhecimento, desenvolveu uma ideia que ajuda a pensar o museu. Em *Science in Action* (1987), Latour (1987) fala em inscrições (*inscriptions*), como gráficos, diagramas, tabelas, mapas e imagens, que tornam os fenômenos móveis, estáveis e combináveis. Um vulcão distante, inacessível, perigoso, é transformado numa série de gráficos de temperatura, pressão, composição química. Esses gráficos podem ser levados para um laboratório em Paris, comparados com gráficos de outros vulcões, sobrepostos, analisados. O vulcão torna-se cognoscível porque foi transformado em inscrição.

A noção de inscrição de Latour (1987) é relevante para a museologia porque evidencia que o ganho cognitivo que o museu proporciona, a capacidade de comparar, de relacionar objetos de épocas e lugares diferentes, tem o custo de o objeto ser transformado, estabilizado, isolado de suas conexões originais. Esse custo é o mesmo que Latour (1987) identifica na produção científica, ao ganhar em combinabilidade e perder em riqueza contextual. O museu é, nessa leitura, um laboratório de inscrições culturais (Latour, 1987, p. 215-257).

Objetos dispersos no tempo e no espaço são coletados, trazidos para um lugar, estabilizados pela conservação, documentados em catálogos, dispostos em sequências, legendados com textos. Eles se tornam móveis (podem ser emprestados, fotografados, reproduzidos), estáveis (podem ser preservados por

décadas ou séculos), combináveis (podem ser postos lado a lado com outros objetos para comparação). O museu transforma o mundo numa coleção cognoscível.

Mas, conforme dissemos, há um preço. O objeto, ao ser musealizado, perde o seu contexto original. A máscara na vitrine já não é a máscara na dança noturna. O fóssil no pedestal já não é o fóssil na rocha, com todas as suas relações geológicas. Há um ganho em cognoscibilidade, mas há uma perda em conexão vital. O objeto torna-se mais fácil de estudar, mas mais difícil de compreender no seu contexto vivo.

Essa tensão entre cognoscibilidade e conexão vital é o que os estudos de patrimônio chamam de dilema da musealização; o mesmo processo que torna o objeto preservável e estudável é o que o aliena de seu contexto original de significação. Não há solução que elimine esse dilema; o que há são estratégias que tentam mitigá-lo, a reconstituição do contexto através de cenografias, a inclusão de vozes das comunidades de origem, o uso de tecnologias digitais para recriar ambientes históricos. Cada estratégia tem seus limites e seus pressupostos.

O educador David Carr (2003), em *The Promise of Cultural Institutions*, propõe uma visão diferente. Para ele, o museu não é um lugar onde se recebe conhecimento pronto, mas um lugar onde se fazem perguntas. Os objetos provocam curiosidade, suscitam questões, estimulam a imaginação. O visitante não é um receptáculo passivo de informação, mas um agente ativo na construção do seu próprio conhecimento. O museu não dá respostas; dá ocasiões para perguntar.

Essa ideia é atraente porque devolve ao visitante um papel ativo. Em vez de ser um mero espectador, ele é um explorador, um investigador, um construtor de sentido. Mas será que qualquer pergunta vale? Será que qualquer interpretação é igualmente válida? Se um visitante vê o fóssil e conclui que é a prova de que os dinossauros viveram com os humanos, isso é tão aceitável quanto a interpretação científica? A maioria dos museus diria que não. Há fatos, há evidências, há métodos, e ignorá-los não é construir conhecimento, é ignorar conhecimento.

George Hein (1998), em *Learning in the Museum*, tenta equilibrar as coisas. Ele fala numa aprendizagem que é construtiva (o visitante constrói o seu próprio

entendimento), mas que se dá na interação entre o que o visitante traz (as suas experiências, os seus preconceitos, as suas perguntas) e o que a instituição oferece (os objetos, os textos, as curadorias, o conhecimento especializado). Nem o visitante tem a palavra final, nem o curador. O conhecimento emerge da conversa entre os dois.

A posição de Hein (1998) pode ser articulada com a epistemologia social de Dewey (1938/1986). O conhecimento é sempre o resultado de uma transação entre um agente e um ambiente, e não a contemplação passiva de um objeto preexistente. No museu, o ambiente é cuidadosamente construído para provocar certas transações, certos atos de percepção, atenção, comparação, inferência. O design da exposição é, nesse sentido, um design epistêmico, por moldar as condições sob as quais o conhecimento pode emergir, sem determiná-lo completamente (Dewey, 1938/1986, p. 25-40).

Essa conversa, porém, nem sempre é fácil. Muitas vezes, os visitantes vêm de culturas, classes sociais, formações educacionais muito diferentes dos curadores. A linguagem dos textos de parede pode ser hermética, a disposição dos objetos pode ser confusa, as referências podem ser estranhas. O museu pode ser um lugar intimidante, que faz o visitante sentir-se ignorante, deslocado, excluído. Aí a conversa não acontece; o que acontece é o silêncio.

Isso nos leva à questão delicada de quem tem autoridade para falar sobre os objetos. Durante séculos, os museus ocidentais falaram sobre as culturas dos outros sem dar voz a esses outros. Exibiram objetos africanos, asiáticos, indígenas, contando histórias que eram as suas histórias, não as histórias dos povos que os produziram. James Clifford (1997), em *Routes: Travel and Translation in the Late Twentieth Century* (1997), chama os museus de zonas de contato (*contact zones*), espaços onde diferentes culturas se encontram, mas em relações de poder profundamente assimétricas. Uma cultura olha, classifica, exibe; a outra é olhada, classificada, exibida.

O conceito de zona de contato, que Clifford (1997) retoma de Mary Louise Pratt, é epistemicamente importante porque descreve uma situação em que o conhecimento não é apenas distribuído desigualmente, mas produzido de modo a reproduzir essa desigualdade. O museu colonial não apenas ignora as perspectivas das culturas representadas; ele produz ativamente um conhecimento

sobre elas que as posiciona como passado, como exotismo, como objetos, e não como sujeitos capazes de falar sobre si mesmas.

A crítica descolonial tem mostrado como essa assimetria é problemática. Linda Tuhiwai Smith (1999), intelectual maori, em *Decolonizing Methodologies*, denuncia como o conhecimento ocidental sempre tratou os povos indígenas como objetos de estudo, não como sujeitos de conhecimento. No museu, isso se traduz em objetos que falam pelos outros, em narrativas que silenciam as vozes locais, em curadorias que ignoram as interpretações das comunidades de origem.

O que fazer? Uma resposta tem sido a curadoria colaborativa, que envolve as comunidades na seleção, na interpretação, na exibição dos objetos que lhes dizem respeito, como parceiros ativos em todo o processo. Isso não resolve todas as assimetrias, mas é um começo. Pelo menos, outras vozes passam a ser ouvidas. A curadoria colaborativa tem uma dimensão epistêmica para além da política, pois, ao incorporar as interpretações das comunidades de origem, o museu enriquece o conhecimento disponível sobre os objetos, corrigindo distorções que resultam do viés do olhar externo. O conhecimento produzido colaborativamente não é apenas mais justo; é também mais completo, pois integra perspectivas que o olhar curatorial ocidental sistemicamente excluía.

4. Ética da Museologia

Chegamos às questões mais difíceis, aquelas que envolvem valores em conflito, direitos em disputa, feridas abertas pela história. Os museus não são instituições neutras. Eles foram, e em grande medida ainda são, lugares de poder. Decidem o que merece ser preservado, o que merece ser mostrado, o que merece ser lembrado. E essas decisões têm consequências éticas profundas.

A ética da museologia não é uma subdisciplina marginal; ela toca em problemas fundamentais de filosofia moral e política, como a legitimidade da propriedade adquirida em condições de violência ou assimetria, os direitos culturais de comunidades sobre seus patrimônios, as obrigações das instituições para com as gerações passadas e futuras, e a justiça distributiva no acesso ao conhecimento e à memória coletiva.

Comecemos pela questão mais candente, que seria a repatriação de objetos. Os museus europeus e norte-americanos estão cheios de objetos que foram retirados dos seus locais de origem em contextos coloniais. Os Mármores de Elgin, no Museu Britânico, foram levados da Grécia no século XIX. Os bronzes do Benin estão espalhados por dezenas de museus no Ocidente, saqueados numa expedição punitiva britânica em 1897. Objetos sagrados de povos indígenas repousam em vitrines de museus de antropologia, longe das comunidades que os veneravam. Restos humanos de ancestrais indígenas estão guardados em gavetas de museus de história natural, como se fossem espécimes de estudo. A pergunta ética em causa aqui seria se esses objetos devem ser devolvidos.

De um lado, há argumentos fortes a favor da repatriação. Os objetos foram levados ilegitimamente, muitas vezes com violência. Pertencem, por direito moral e histórico, às comunidades de origem. Fazem parte da identidade desses povos, da sua memória, da sua cultura. A sua presença em museus estrangeiros é um lembrete constante da espoliação colonial, uma ferida aberta que não cicatriza. Devolvê-los é um ato de reparação, de justiça histórica. A Grécia reivindica os Mármores de Elgin há décadas; a Nigéria luta pela devolução dos bronzes do Benin; povos indígenas de todo o mundo exigem a volta dos seus ancestrais.

Do ponto de vista da teoria da justiça, a reivindicação de repatriação pode ser fundamentada em pelo menos no princípio da reparação (que exige a correção de danos históricos) e no princípio dos direitos culturais (que reconhece que comunidades têm direitos especiais sobre os objetos que constituem a sua identidade e memória coletiva). Esses dois princípios podem convergir ou divergir conforme o caso. Objetos adquiridos por compra, ainda que em condições de assimetria de poder, têm uma história moral diferente dos objetos saqueados, e a ética da repatriação precisa ser sensível a essas diferenças.

O arqueólogo e museólogo James Cuno (2008), no entanto, defende uma posição contrária. Em *Who Owns Antiquity?* (2008), conforme dissemos e agora mais aprofundadamente, Cuno (2008) argumenta que os chamados museus universais ou enciclopédicos, como o Louvre, o Museu Britânico, o Metropolitan, cumprem uma função moral importante. Eles tornam acessível a um público global obras de arte e artefatos de todas as culturas e períodos, promovendo a

compreensão intercultural, o apreço pela diversidade, o estudo comparado. Repatriar esses objetos, segundo Cuno (2008), seria fragmentar esse patrimônio comum, dispersar coleções formadas ao longo de séculos, e reduzir o acesso global a eles, confinando as obras aos seus países de origem, onde muitas vezes não há a mesma infraestrutura de preservação, o mesmo público, a mesma visibilidade.

A posição de Cuno (2008) é atraente para muitos, mas enfrenta críticas sérias. A primeira é que o discurso da "humanidade" e do "patrimônio comum" muitas vezes mascara certas relações de poder. Quem decide o que é patrimônio da humanidade? Quem decide o que deve ser preservado e como? Na prática, são os curadores europeus e norte-americanos, a partir de perspectivas ocidentais. A "humanidade" que decide é, na verdade, um grupo muito específico de pessoas em posições de poder. Além disso, muitos objetos têm significados espirituais, cerimoniais e culturais específicos para as comunidades de origem. A máscara não é apenas uma obra de arte; é um objeto sagrado, que só devia ser vista em contextos rituais. O crânio não é apenas um espécime científico; é um ancestral, que merece descanso digno. Tratá-los como meras peças de museu é uma violência simbólica que se acrescenta às violências materiais do passado.

A crítica ao universalismo de Cuno (2008) pode ser articulada a partir da distinção entre valores universais e perspectivas particulares. O valor da preservação e do acesso ao conhecimento é, de fato, universal; mas a forma específica assumida pelo museu enciclopédico ocidental não é a única, nem necessariamente a melhor, forma de realizar esses valores. O argumento universalista pode ser usado tanto para justificar a repatriação (os objetos pertencem à humanidade, não a museus específicos) quanto para resistir a ela, dependendo de como se concebe o acesso universal ao patrimônio comum.

O argumento do acesso global também é problemático. O acesso aos grandes museus dos EUA e da Europa é, na prática, restrito a uma minoria privilegiada, a saber, quem pode viajar, quem tem passaporte, quem tem dinheiro. Para a enorme maioria da população mundial, especialmente nos países de origem dos objetos, esses museus são tão inacessíveis como se estivessem em outro planeta.

Uma posição intermediária, defendida por Kwame Anthony Appiah (2006) em *Cosmopolitanism*, reconhece tanto o valor do acesso global quanto os direitos das comunidades. Appiah (2006) propõe um cosmopolitismo enraizado (*rooted cosmopolitanism*), no qual temos obrigações para com a humanidade como um todo, mas também obrigações especiais para com as comunidades a que pertencemos. O cosmopolitismo enraizado de Appiah (2006) pode ser aproximado do princípio da subsidiariedade, pois as decisões sobre os objetos devem ser tomadas no nível mais próximo possível das comunidades afetadas, exceto quando há razões específicas para envolver níveis mais amplos de governança. Aplicado à museologia, isso sugere uma presunção a favor das comunidades de origem, que pode ser afastada por razões específicas, como a incapacidade demonstrada de preservação ou o risco de destruição num contexto de conflito, mas não pela mera preferência das instituições custodiadoras.

Há também a questão da autoridade interpretativa. Quem tem o direito de contar a história dos objetos? Durante muito tempo, essa autoridade coube aos curadores, aos especialistas. Mas as comunidades representadas nas coleções têm cada vez mais reivindicado o direito de participar da interpretação do seu próprio património. O antropólogo Michael Ames (1992), em *Cannibal Tours and Glass Boxes*, mostra como os museus antropológicos representaram as culturas não-ocidentais de forma distorcida, exotizante, primitivizante. Os objetos eram apresentados como testemunhos de um passado distante, e os povos contemporâneos, quando apareciam, eram tratados como meros sobreviventes, curiosidades. Essa forma de representação não é apenas epistemicamente inadequada; é moralmente problemática, pois contribui para a perpetuação de estereótipos, para a desvalorização das culturas representadas, para a manutenção de relações de poder coloniais.

A filósofa Miranda Fricker (2007), em *Epistemic Injustice*, oferece ferramentas para pensar essas questões. As duas formas de injustiça epistêmica identificadas por Fricker (2007) têm aplicações específicas no contexto museal que vale distinguir. A injustiça testemunhal pode ocorrer mesmo quando a comunidade tem voz no processo curatorial, se suas interpretações forem sistematicamente subordinadas ou enquadradas pelas categorias do curador externo. A injustiça hermenêutica é mais profunda; ela ocorre quando as próprias categorias do discurso museal são incompatíveis com os modos de compreender e

articular o mundo das comunidades representadas, quando, por exemplo, a distinção ocidental entre “sagrado” e “profano”, ou entre “arte” e “artefato”, não encontra correspondência nas categorias locais.

A saída, mais uma vez, passa por práticas de curadoria colaborativa, em que as comunidades são envolvidas como parceiras ativas, não como meras informantes. Vários museus têm experimentado esse modelo, com resultados encorajadores. O Museu do Quai Branly, em Paris, por exemplo, tem trabalhado com comunidades de origem para repensar a exibição dos seus acervos. O Museu Nacional do Índio Americano, em Washington, foi concebido desde o início com participação indígena. Ainda há muito a fazer, mas o caminho está sendo aberto.

Há também a tensão entre preservação e acesso. Os museus têm, por um lado, a missão de preservar os objetos para as gerações futuras. Isso implica medidas restritivas, como controlo de temperatura e humidade, iluminação reduzida, manuseio limitado. Por outro lado, têm a missão de proporcionar acesso público aos objetos. Um objeto que nunca é visto cumpre apenas metade da sua função. O equilíbrio entre esses valores é delicado. Obras-primas da pintura não podem ser expostas permanentemente, sob pena de deterioração irreversível. Objetos extremamente frágeis, como tecidos antigos, podem ser expostos apenas em condições muito controladas. A digitalização tem aberto novas possibilidades, permitindo o acesso virtual a objetos que não podem ser expostos fisicamente. Mas a experiência digital não substitui completamente a experiência presencial. Há algo em estar diante do objeto original, em sentir a sua presença, que a reprodução não capta.

Esse “algo” que a presença do objeto original oferece e que a reprodução não capta foi descrito por Walter Benjamin (1936/2012) como aura, a aparição única de uma distância, por mais próxima que esteja o objeto. Para Benjamin (1936/2012), a reprodutibilidade técnica destrói a aura ao multiplicar as cópias e descontextualizar o original. No museu, porém, a situação é mais complexa. O objeto original está fisicamente presente, mas foi deslocado de seu contexto aurático originário (o ritual, o culto, o uso). A aura museal é, portanto, uma aura deslocada, que Benjamin (1936/2012) chamaria de valor de exposição em contraposição ao valor cultual (Benjamin, 1936/2012, p. 179-212).

Há ainda a questão do acesso democrático. Os museus tradicionais sempre foram, em certa medida, instituições excludentes. A arquitetura imponente, inspirada em templos clássicos, pode intimidar visitantes de classes populares. A linguagem dos textos de parede, muitas vezes especializada, pode ser inacessível para pessoas com baixa escolaridade. Os horários de funcionamento, geralmente diurnos, podem excluir quem trabalha durante o dia. A localização, frequentemente nos centros das grandes cidades, pode dificultar o acesso de quem mora na periferia.

A historiadora da arte Carol Duncan (1995), em *Civilizing Rituals*, argumenta que os museus de arte funcionam como espaços de ritual civilizatório. Ao entrar no museu, o visitante é convidado a participar de um ritual que o separa do mundo exterior e o insere num espaço sagrado, onde deve comportar-se de determinada maneira. Esse ritual não é neutro; ele comunica mensagens sobre quem pertence e quem não pertence, sobre quem é capaz de apreciar a "alta cultura" e quem é excluído dela. Democratizar o acesso aos museus, portanto, não é apenas uma questão de eliminar barreiras financeiras. É também uma questão de transformar a cultura institucional, de tornar os museus mais acolhedores para públicos diversos, de diversificar a linguagem, de envolver as comunidades na programação, de levar exposições para fora dos muros do museu.

A análise de Duncan (1995) pode ser conectada à teoria do capital cultural de Pierre Bourdieu (1969/1991), no qual o museu de arte funciona como um espaço em que o capital cultural das classes dominantes é naturalizado como gosto universal. Quem possui o *habitus* adequado, os esquemas de percepção e apreciação associados à educação artística burguesa, experimenta o museu como um espaço de confirmação e prazer; quem não o possui experimenta-o como um espaço de exclusão e incompetência. A democratização do museu requer, portanto, uma crítica reflexiva dos pressupostos de classe que estruturam tanto a curadoria quanto a arquitetura museal (Bourdieu; Darbel, 1969/1991, p. 37-54).

Uma questão particularmente delicada é a exibição de restos humanos. Museus de antropologia, história natural, arqueologia e medicina frequentemente possuem coleções de restos humanos, com esqueletos, múmias, crânios, partes do corpo preservadas. Esses restos foram, na sua maioria, adquiridos em contextos moralmente problemáticos, como escavações de cemitérios indígenas

sem consentimento, coleta em campos de batalha, remoção de corpos de pessoas escravizadas, dissecação de corpos de pobres não reclamados. Para muitas comunidades, especialmente as indígenas, esses restos são ancestrais, pessoas que merecem respeito e descanso digno, não objetos de exibição ou estudo. A aprovação de leis como a *Native American Graves Protection and Repatriation Act* (NAGPRA) nos Estados Unidos, em 1990, marcou um ponto de virada. A lei reconheceu o direito de comunidades indígenas de reivindicar a repatriação de restos humanos e objetos funerários dos seus ancestrais. Desde então, milhares de conjuntos de restos foram devolvidos e enterrados adequadamente.

A ética da exibição e do estudo de restos humanos pode ser enquadrada a partir do princípio do respeito às pessoas, formulado por Kant (1785/2005) como a exigência de tratar as pessoas sempre como fins em si mesmas, nunca apenas como meios. Os restos humanos são, em certo sentido, "restos de pessoas", mantendo uma conexão com a dignidade dos indivíduos e das comunidades a que pertenceram. Tratá-los como espécimes científicos sem o consentimento das comunidades descendentes é violar esse princípio, independentemente do valor do conhecimento que possa ser obtido (Kant, 1785/2009, p. 66-67). No entanto, pesquisadores argumentam que o estudo de restos humanos pode fornecer informações valiosas sobre a saúde, a nutrição, as doenças, as migrações de populações do passado. Como equilibrar o respeito pelos mortos com o valor do conhecimento científico? Mais uma vez, não há resposta fácil. O que parece essencial é que as decisões sejam tomadas através de diálogo genuíno com as comunidades afetadas, e não impostas unilateralmente.

O princípio do consentimento informado, fundamental na ética da pesquisa com seres humanos, pode ser adaptado para o contexto dos restos humanos históricos, ao defendermos que as comunidades descendentes devem ser reconhecidas como as partes com autoridade para consentir ou recusar o estudo e a exibição dos restos de seus ancestrais. Esse reconhecimento não elimina os conflitos, mas os deslocada da oposição entre ciência e tradição para a negociação entre interesses legítimos de partes com estatutos diferentes.

5. Considerações Finais

Os museus são lugares estranhos e fascinantes. Guardam objetos que já não são o que eram, que foram deslocados no tempo e no espaço, que ganharam novas vidas, novos significados, novas funções. São lugares de memória, mas também de esquecimento, porque escolher o que guardar é também escolher o que descartar. São lugares de conhecimento, mas também de poder, porque decidir o que mostrar e como mostrar é decidir que histórias serão contadas e quais serão silenciadas.

Essa duplicidade, memória e esquecimento, conhecimento e poder, não é uma imperfeição contingente dos museus que poderia ser eliminada por reformas adequadas. É uma característica estrutural de qualquer instituição que seleciona, classifica e exibe, pois a seleção é sempre também exclusão, a classificação é sempre também hierarquização, a exibição é sempre também enquadramento. O desafio não é eliminar essa duplicidade, mas torná-la explícita e submetê-la à crítica reflexiva.

A metafísica da museologia nos mostrou que os objetos museais têm uma ontologia complexa, híbrida, dual. São matéria e significado, coisa e signo, original e cópia, individual e coletivo. A sua identidade não é fixa; ela é negociada, construída, disputada. A restauração, a conservação, a exibição, todas essas práticas intervêm no objeto, transformam-no, acrescentam-lhe camadas. O objeto de museu é um ser histórico, que carrega em si as marcas do tempo e das intervenções humanas.

Essa historicidade do objeto museal tem um corolário metodológico, a saber, a análise de um objeto museal não pode se limitar à análise de suas propriedades físicas atuais; ela deve reconstituir a história das intervenções que o produziram como objeto museal, incluindo a aquisição, o transporte, a restauração, a exibição, a reinterpretação. O objeto museal é, em um sentido preciso, um palimpsesto, i.e., um texto em que camadas sucessivas de inscrição se sobrepõem sem que as anteriores desapareçam completamente.

A epistemologia nos mostrou que o conhecimento museal é construído, não descoberto. As exposições são interpretações, narrativas, escolhas. Os objetos não falam por si mesmos; são os curadores que falam por eles, e os visitantes que ouvem (ou não) o que é dito. Essa mediação é inevitável, mas não é neutra.

Ela carrega valores, pressupostos, relações de poder. O desafio é tornar essa mediação mais consciente, mais reflexiva, mais aberta ao diálogo com diferentes vozes.

A ética nos confrontou com os dilemas mais difíceis. A repatriação de objetos, a autoridade interpretativa, o equilíbrio entre preservação e acesso, o tratamento dos restos humanos, todas essas questões envolvem valores em conflito, direitos em disputa, feridas históricas. Não há respostas prontas, não há fórmulas mágicas. O que há são os processos de diálogo, negociação, escuta, respeito. E os princípios de justiça, reparação, dignidade, humildade. Todo o mais é uma discussão sobre como ponderar esses princípios e usar os processos para realizar nossa atividade no dia a dia.

A ausência de respostas prontas não implica que todas as posições sejam igualmente válidas. Implica, antes, que a validade das posições deve ser avaliada através de processos deliberativos que incluam todas as partes afetadas, e não decidida unilateralmente pelas instituições com maior poder. A filosofia da museologia contribui para esses processos ao clarificar os pressupostos que estruturam as disputas, identificar os valores em jogo e propor distinções que permitem avançar onde o senso comum se confunde.

O pluralismo que emerge dessas reflexões não é uma posição teórica abstrata; é uma constatação prática. Diferentes objetos, diferentes contextos, diferentes comunidades exigem diferentes respostas. O que serve para um museu de arte pode não servir para um museu de história natural. O que funciona na Europa pode não funcionar na África. A tarefa de quem pensa os museus, de quem trabalha nos museus, de quem visita os museus, é estar atento a essas diferenças, a essas nuances, a esses *trade-offs* entre teorias.

Esse pluralismo contextual é distinto do relativismo, pois este renunciaria a qualquer padrão normativo. Aquele pressupõe a existência de princípios gerais, justiça, respeito, transparência, diálogo, e reconhece que a aplicação desses princípios a situações específicas requer sensibilidade às particularidades de cada caso. É o que Aristóteles (1973) chamaria de *phronesis*, a virtude do julgamento prático que adapta os princípios gerais às circunstâncias particulares sem dissolvê-los em puro oportunismo.

Os museus são, no fundo, uma das formas mais importantes que a humanidade inventou para lidar com o tempo, com a memória, com a morte. Guarda-

mos objetos para que eles sobrevivam a nós, para que as gerações futuras possam ver o que fomos, o que fizemos, o que valorizamos. Nessa tarefa, erramos, acertamos, aprendemos. O importante é continuar pensando, continuar questionando, continuar buscando formas melhores de lembrar e de esquecer.

Essa tarefa não é apenas institucional; é também filosófica. A filosofia da museologia não termina na análise das práticas existentes; ela aponta para uma prática reflexiva, a prática de criar, gerir e visitar museus com consciência das implicações ontológicas, epistemológicas e éticas de cada decisão. O museu não é apenas um arquivo do passado; é também uma proposição sobre o futuro, sobre o que vale preservar, sobre quem tem direito à memória, sobre que tipo de conhecimento queremos construir.

9

Filosofia do Turismo

1. Por que uma Filosofia do Turismo?

Pensa numa praia. Qualquer praia. Durante a semana, fora de temporada, ela está quase vazia. Alguns pescadores puxam as redes, umas poucas crianças brincam na beira d'água, um casal caminha de mãos dadas. O vento sopra, as ondas quebram, o sol esquenta. É só uma praia. Aí chega o fim de semana. Depois, o feriado. Depois, as férias de verão. A praia se enche. Barracas, cadeiras, guarda-sóis, vendedores ambulantes, música, gritos, risadas. Gente de fora, com sotaque diferente, com hábitos diferentes, com dinheiro diferente. Os pescadores ainda estão lá, mas agora eles também vendem peixe frito para os turistas. As crianças da terra dividem a beira com crianças que vieram de longe. O casal que caminhava agora desvia de grupos que tiram selfies.

A praia mudou? A areia é a mesma, o mar é o mesmo, o vento é o mesmo. Mas algo mudou, e mudou profundamente. A praia deixou de ser apenas um lugar e passou a ser um destino. Ganhou uma nova camada de existência. Agora ela existe não só para quem vive ali, mas também para quem vem de fora. Ela é produto, é imagem, é experiência, é mercadoria.

Essa transformação, de lugar a destino, é o ponto de entrada da filosofia do turismo. O que acontece com um lugar quando ele adquire uma segunda existência, voltada para o olhar e o consumo externos? A pergunta não é apenas sociológica; é também metafísica, pois implica uma mudança na estrutura de sentido do próprio lugar, que passa a existir para além de sua função de habitação originária.

O que aconteceu com a praia é o que acontece com milhares de lugares no mundo todo, todos os dias. O turismo transforma lugares. Às vezes para melhor, às vezes para pior, quase sempre de formas ambivalentes. Traz dinheiro, em-

prego, desenvolvimento. Mas também traz inflação, especulação imobiliária, descaracterização, conflitos. As pessoas que vivem nos lugares turísticos muitas vezes se perguntam se a cidade ainda é sua, se a praia ainda é sua. Ou o nativo se tornou um estranho na sua própria terra?

E os turistas? O que eles estão buscando? Muitos dizem que querem "autenticidade". Querem conhecer a "cultura local", experimentar a "comida típica", ver as "tradições verdadeiras". Mas o que é essa tal autenticidade? Será que ela existe mesmo? Ou será que os turistas buscam uma coisa que, quando encontram, já não é mais a mesma porque eles próprios a transformaram?

A tensão entre a busca de autenticidade e a sua inevitável alteração pela presença do próprio turista constitui o paradoxo central da filosofia do turismo, pois o objeto da busca é perturbado pelo ato de buscá-lo. Esse paradoxo tem uma estrutura lógica análoga ao princípio da incerteza em física, em que o ato de medir interfere no estado do sistema medido, e será o fio condutor das seções seguintes. Pense no viajante que vai a uma aldeia remota buscar a "cultura autêntica". Quando ele chega, a sua presença já altera a aldeia. As crianças pedem moedas, os adultos oferecem artesanato feito para turistas, os velhos contam histórias adaptadas para o visitante. A autenticidade que ele buscava, se é que existia, desfez-se no momento em que ele a procurou. O turista quer ver a cultura como ela é, mas a sua própria presença faz com que ela deixe de ser como era.

Há também a questão do conhecimento. Viajar amplia os horizontes, todo mundo diz. A gente aprende sobre outros lugares, outras culturas, outras formas de vida. Mas será que é bem assim? Será que a gente realmente aprende, ou será que a gente apenas confirma os nossos próprios preconceitos, vê o que esperávamos ver, fotografa o que já conhecíamos dos filmes e das revistas? Quantas vezes um turista volta de viagem e diz "que lugar incrível, as pessoas são tão felizes, tão simples"? O que ele sabe, de fato, sobre a felicidade ou a simplicidade daquelas pessoas, que ele viu por alguns dias, de fora, sem falar a língua, sem partilhar a vida?

E depois há as questões éticas, as mais espinhosas. O turista tem o direito de ir a qualquer lugar? As comunidades têm o direito de fechar as portas? Cidades como Barcelona, Veneza, Amsterdã estão a dizer "chega". Moradores vão para as ruas protestar contra a massificação, contra a transformação dos seus

bairros em parques temáticos, contra a impossibilidade de viver nas suas próprias casas. O que fazer quando o direito de viajar colide com o direito de viver em paz?

Essa colisão não é meramente econômica; é, em seu núcleo, um conflito entre direitos reconhecidos no plano internacional. O direito à mobilidade e ao lazer, inscrito no artigo 24 da Declaração Universal dos Direitos Humanos, choca-se com o direito à habitação digna, ao espaço público e à identidade cultural das comunidades receptoras. A filosofia do turismo precisa articular esses direitos em conflito, e não apenas descrevê-los.

O turismo é uma indústria gigantesca, um dos maiores negócios do planeta. Move multidões, movimenta fortunas, transforma paisagens, destrói ecossistemas, constrói impérios. Mas também é uma experiência profundamente pessoal, íntima até. As memórias de uma viagem podem marcar uma vida inteira. Um pôr do sol visto num lugar distante pode nos acompanhar para sempre.

Este capítulo é uma tentativa de pensar sobre tudo isso. Vamos começar pela metafísica, pelo que acontece quando um lugar se torna turístico. Depois vamos para a epistemologia, para o que a gente aprende (ou não) viajando. E vamos terminar na ética, com as perguntas sobre hospitalidade, justiça e responsabilidade. É uma conversa, não uma aula. E como toda boa conversa, não tem destino certo; o que importa é o caminho. A articulação dessas três dimensões não é meramente expositiva. Elas se implicam mutuamente. A resposta que se dá à questão metafísica sobre o que é um lugar turístico determina o que contará como conhecimento genuíno sobre ele, o que por sua vez delimita as obrigações éticas de quem o visita e de quem o habita.

2. Metafísica do Turismo

Voltemos à praia. Durante a semana, fora de temporada, é só uma praia. Lugar de trabalho para os pescadores, de brincadeira para as crianças, de passeio para o casal. É um lugar que existe para quem vive ali. As relações são diretas, práticas, vividas. O pescador conhece o mar, sabe onde os peixes estão, sabe quando vai chover. A criança conhece as pedras, sabe onde tem ouriço, sabe onde a água é mais funda. O casal conhece a trilha, sabe onde o vento bate menos, sabe onde o pôr do sol é mais bonito.

Essa relação com o lugar é o que Martin Heidegger (1954/2008) chamaria de habitar (*Wohnen*), i.e., uma forma de ser-no-mundo anterior a qualquer representação ou consumo, em que o lugar é dado como horizonte vivido de prática e sentido, não como objeto diante de um sujeito. O turismo interrompe esse habitar ao introduzir uma relação de exterioridade; o lugar passa a existir para quem não o habita, adquirindo a estrutura do espetáculo antes mesmo de tornar-se mercadoria (Heidegger, 1954/2008, p. 125-130).

Aí chega o turismo. No começo, são poucos. Aventureiros, mochileiros, gente que busca lugares diferentes. Eles conversam com os pescadores, compram peixe, dormem em redes. Ainda são bem-vindos. Depois, mais gente começa a vir. Aparecem pousadas, restaurantes, lojas de souvenir. Os pescadores já não pescam tanto; alugam barcos para passeios. As crianças aprendem a vender artesanato. O casal abre uma pousada.

A praia agora é outra. Não é que a areia tenha mudado, ou o mar, ou o vento. Mas a praia ganhou uma nova função, um novo significado. Ela não é mais apenas lugar de vida; é também lugar de consumo. Existe para quem vive ali, sim, mas existe também para quem vem de fora. Essa dupla existência é frágil, tensa. Muitas vezes, uma acaba engolindo a outra.

O geógrafo Tuan (1977) teorizou essa tensão através da distinção entre espaço (*space*) e lugar (*place*). O espaço é abstrato, indiferenciado, aberto ao movimento; o lugar é concreto, investido de valor, habitado. O turismo opera uma conversão do lugar em espaço, desmanchando os significados sedimentados pela habitação e submetendo o território ao fluxo impessoal do deslocamento (Tuan, 1977, p. 54-61).

O sociólogo americano MacCannell (1976), nos anos 1970, tentou entender essa transformação. Ele partiu da ideia simples de que a vida social é feita de performances. A gente desempenha papéis. No trabalho, a gente é profissional; em casa, a gente é pai, mãe, filho; com os amigos, a gente é descontraído. Há espaços onde a gente atua para o público, que MacCannell chama de palco (*front stage*), e espaços onde a gente relaxa, onde cada um é "si mesmo", os bastidores (*back stage*).

MacCannell recorre explicitamente a Erving Goffman (1959), que havia desenvolvido a distinção palco/bastidores em sua análise da apresentação do eu na vida cotidiana. Para Goffman (1959), toda interação social envolve uma ges-

tão de impressões, a saber, o ator gerencia o que o público vê, reservando para os bastidores os comportamentos que não quer expor. MacCannell transpõe essa microanálise para a escala do turismo, mostrando que a própria organização espacial dos destinos reproduz a lógica goffmaniana (Goffman, 1959, p. 22-30; Maccannell, 1976, p. 91-102).

O turista, segundo MacCannell, busca os bastidores. Quer ver a "vida real", a cultura "autêntica", o lugar como ele é "de verdade". Não quer o palco preparado para ele; quer o que está por trás. O problema é que a indústria do turismo aprendeu a transformar bastidores em palco. Cria uma autenticidade encenada (*staged authenticity*). Mostra uma versão limpa, organizada, segura, palatável da vida local, que parece bastidor mas é, na verdade, palco.

O conceito de autenticidade encenada dialoga com a crítica de Guy Debord (1967/1997) à sociedade do espetáculo, de que numa sociedade em que a produção de imagens domina a produção de coisas, a experiência autêntica é sistematicamente substituída por sua representação. O turismo seria, nessa leitura, uma das formas mais acabadas da lógica do espetáculo, em que o lugar é consumido como imagem antes de ser vivido como realidade (Debord, 1967/1997, p. 13-20).

O ritual religioso que o turista assiste foi encurtado, simplificado, despojado dos seus elementos mais sagrados ou perturbadores. A técnica artesanal que ele vê foi adaptada para produzir mais rápido, para agradar ao gosto estrangeiro. A conversa que ele tem com o morador foi ensaiada, treinada, padronizada. O turista pensa que viu os bastidores, mas viu apenas mais um palco, preparado especialmente para ele. MacCannell é pessimista. Para ele, o turista está preso num jogo de espelhos, sempre buscando uma autenticidade que lhe escapa, sempre encontrando apenas representações. Quanto mais ele busca, mais a indústria produz representações. É uma espécie de armadilha.

Mas será que é sempre assim? O antropólogo Edward Bruner (2005) desconfia. Ele estudou o turismo na Indonésia, na África, e viu coisas mais complicadas. Muitas das tradições que os turistas vêm ver não são "autênticas" num sentido antigo; foram criadas recentemente, às vezes com a participação dos próprios turistas. Uma dança que parece tradicional pode ter sido inventada há vinte anos para um festival de turismo. Isso a torna falsa? Ou é apenas uma tradição nova, uma cultura viva que se reinventa?

Bruner (2005) conecta essa observação à análise de Eric Hobsbawm e Terence Ranger (1983) sobre a invenção de tradições. Muitas práticas que se apresentam como imemorialmente antigas foram deliberadamente criadas ou formalizadas em períodos recentes, em resposta a necessidades políticas ou econômicas do presente. No contexto do turismo, isso significa que a tradição autêntica pode ser, simultaneamente, inventada e genuinamente vivida pela comunidade que a pratica (Hobsbawm; Ranger, 1983, p. 1-14; Bruner, 2005, p. 17-30).

Além disso, Bruner (2005) viu que o turismo pode ter efeitos positivos. Pode dar valor a práticas culturais que estavam em risco de desaparecer. Pode gerar orgulho nas comunidades, que passam a ver a sua cultura como algo digno de ser mostrado. Pode financiar a preservação de património. O turismo não é só destruição; é também criação. E se atentarmos bem, essa exposição do patrimônio preservado se assemelha bastante ao processo de se tornar um semióforo que o objeto museal costuma passar.

Esse efeito de revalorização cultural pelo olhar externo tem uma lógica paradoxal, pois a mesma dinâmica que pode alienar a cultura local, transformando-a em produto, pode também mobilizar recursos para sua preservação e difusão. Esse paradoxo foi investigado por Barbara Kirshenblatt-Gimblett (1998) no conceito de patrimônio como modo de produção cultural (*heritage as a mode of cultural production*). O passado é reativado para o consumo presente sem necessariamente ser destruído, e a exibição pública confere ao objeto cultural uma segunda existência análoga à do objeto museificado, o que o texto original intui ao mencionar o processo de tornar-se semióforo (Kirshenblatt-Gimblett, 1998, p. 149-165).

O sociólogo britânico Urry (1990) propôs uma maneira diferente de pensar. Para ele, o que importa não é tanto o que os lugares são, mas como eles são vistos. Urry (1990) fala em olhares turísticos (*tourist gazes*). O turista não vê o lugar como ele é; vê através de lentes que aprendeu nos guias, nos filmes, nas revistas, nas redes sociais. Essas lentes moldam o que ele vê, o que ele acha importante, o que ele fotografa.

O conceito de olhar turístico foi desenvolvido por Urry (1990) em diálogo com Michel Foucault (1975); e a ideia é que assim como o olhar médico (*regard médical*) organiza o espaço clínico de acordo com categorias disciplinares es-

pecíficas, o olhar turístico organiza o espaço do destino de acordo com categorias de prazer, diferença e autenticidade historicamente produzidas. O olhar não é neutro; é um dispositivo de poder que seleciona o que é digno de ser visto e descarta o resto (Urry, 1990, p. 1-15; Foucault, 1963/2003, p. 9-29).

Há diferentes tipos de olhar. O olhar romântico busca a solidão, a natureza intocada, a experiência quase espiritual. O olhar coletivo busca a multidão, a animação, a festa. O olhar antropológico busca a cultura, a diferença, a autenticidade. Cada olhar constrói o lugar de uma maneira diferente. A mesma praia pode ser, para um, paraíso isolado; para outro, point badalado; para outro, amostra de cultura local. A tipologia dos olhares em Urry (1990) foi criticada e ampliada por Jonas Larsen (2001), que argumentou que o olhar turístico não é apenas visual, mas implica práticas corporificadas e afetivas como tocar, cheirar, saborear, mover-se. O turismo seria menos um olhar do que uma performance encarnada (Larsen, 2001, p. 80-98). Essa crítica antecipa o programa da virada performativa nos estudos do turismo, que desloca o foco da representação para a prática.

O que Urry (1990) nos ajuda a ver é que o lugar turístico não é uma entidade fixa. Ele é performado, construído, interpretado, reinterpretado. O pescador, a criança, o casal, o turista, cada um vê a praia de um jeito, cada um vive a praia de um jeito, cada um constrói a praia de um jeito. A praia é todas essas coisas ao mesmo tempo, e a tensão entre elas é o que faz dela um lugar turístico. Essa pluralidade de construções do mesmo lugar pode ser teorizada através do conceito de mundos de sentido sobrepostos, i.e., o mesmo objeto físico é habitado por ontologias práticas distintas que coexistem sem se fundir. O conflito turístico é, em grande parte, um conflito entre mundos de sentido que disputam a primazia sobre um mesmo território.

E a autenticidade nessa história? Talvez ela não seja uma propriedade dos lugares, mas uma qualidade da experiência. Há quem distinga autenticidade objetiva (o objeto é mesmo o que parece ser, tem a origem que diz ter) de autenticidade existencial (a experiência é sentida como genuína, toca fundo, transforma). O turista que se emociona com um pôr do sol numa praia lotada pode estar tendo uma experiência autenticamente transformadora, mesmo que a praia esteja cheia de barracas e vendedores e música alta. A autenticidade não está na praia; está nele.

Essa distinção foi sistematizada por Wang (1999) Ning, que acrescentou uma terceira modalidade de autenticidade, a construtiva, reconhecendo que a autenticidade é parcialmente projetada pelos observadores sobre os objetos. A autenticidade objetiva corresponde a um certo realismo; a construtiva reconhece a negociação intersubjetiva do que conta como autêntico; a existencial refere-se a um estado de ser em que o sujeito está plenamente presente para si mesmo, indiferente à autenticidade dos objetos externos. Essa taxonomia revela que muitos conflitos no debate sobre autenticidade turística são conflitos terminológicos, não substantivos (Wang, 1999, p. 349-370). A experiência turística é mais complexa do que as teorias simplistas supõem. Não há receita para a autenticidade. Há encontros, alguns mais felizes, outros menos.

E o turista? Quem é essa figura tão criticada, tão ridicularizada, tão presente? Há quem distinga viajante de turista. O viajante seria aquele que busca experiências autênticas, que se aventura, que se mistura; o turista seria aquele que consome pacotes, que vê o que mandam ver, que fica na superfície. Mas a distinção é falsa. Todo viajante é turista para alguém. Todo turista, em algum momento, tem uma experiência genuína. As pessoas não são tipos; são momentos, são performances. Hoje posso ser turista, amanhã viajante, e na semana que vem, quem sabe, os dois ao mesmo tempo.

A idealização do viajante em oposição ao turista é frequentemente uma forma de distinção de classe que usa o capital cultural para se diferenciar do turismo de massa. Zygmunt Bauman (1998) observou que a figura do turista e a do vagabundo são dois polos de um continuum de mobilidade e pertencimento, não categorias fixas; o mesmo indivíduo pode ocupar posições diferentes em contextos diferentes, e a oposição valorativa entre ambos serve, antes de tudo, para legitimar privilégios de mobilidade (Bauman, 1998, p. 77-102).

3. Epistemologia do Turismo

Todo mundo diz que viajar amplia os horizontes. Que a gente aprende sobre outras culturas, sobre outros modos de vida, sobre o mundo. Mas será que é verdade? O que a gente realmente aprende quando viaja? Vamos pegar um exemplo concreto. Um turista brasileiro vai a Paris. Ele tira foto na Torre Eiffel, visita o Louvre, vê a Mona Lisa, come um croissant, caminha pelos Champs-

Élysées. O que ele aprendeu sobre Paris? Aprendeu que a Torre Eiffel é grande, que a Mona Lisa é pequena, que os croissants são gostosos, que os Champs-Élysées são cheios de lojas. Aprendeu, talvez, que os parisienses são simpáticos ou são grossos, dependendo do atendimento que teve. Aprendeu alguma coisa sobre a história de Paris, sobre a sua arquitetura, sobre a sua vida cultural, sobre os seus conflitos sociais, sobre as pessoas que lá vivem e não são vistas pelos turistas? Provavelmente não.

Há uma assimetria epistêmica estrutural entre o turista e o habitante. O habitante possui o que Michael Polanyi (1966) chamou de conhecimento tácito (*tacit knowledge*), um saber prático, corporificado, adquirido pela experiência repetida e pela imersão no ambiente. O turista, ao passar rapidamente pelo lugar, só pode acessar o conhecimento explícito, proposicional, tematizado nos guias e nos roteiros. A maioria do que faz Paris ser Paris para quem lá vive, os ritmos cotidianos, as redes de sociabilidade, as tensões de classe e raça, é invisível ao olhar turístico (Polanyi, 1966, p. 4-17).

O crítico literário Jonathan Culler (1981) tem uma tese provocadora. Ele diz que o turista não busca conhecimento novo; busca reconhecimento. Quer ver com os próprios olhos o que já viu mil vezes em imagens. Quer confirmar que a Torre Eiffel é como nos postais, que a Mona Lisa é como nos livros, que Paris é como nos filmes. A experiência turística é, em grande parte, uma experiência de *déjà vu*. A gente viaja para ver o que já sabe, não para descobrir o que não sabe. Essa tese de Culler (1981) dialoga com a noção de horizonte de expectativas (*Erwartungshorizont*) de Hans Robert Jauss (1974/1994), na qual toda experiência estética ou cultural é estruturada por expectativas pré-formadas que determinam como o objeto é recebido. No turismo, o horizonte de expectativas é formado antes da viagem pelas imagens midiáticas do destino, e a experiência real é julgada em relação a esse horizonte. O prazer do turista pode vir precisamente da confirmação do horizonte, não de sua ruptura, o que coloca o turismo de massa em tensão com qualquer pedagogia da novidade (Jauss, 1974/1994, p. 23-32; Culler, 1981, p. 127-140). É por isso que as fotos dos turistas são tão parecidas. Todo mundo fotografa os mesmos lugares, dos mesmos ângulos, na mesma luz. As fotos não são registros de descoberta; são provas de que se esteve lá. “Estive na Torre Eiffel, olha a prova”. A foto é um certificado de presença, não um testemunho de conhecimento.

A análise semiótica das fotografias turísticas foi aprofundada por Roland Barthes (1980/1984), que distinguiu o *studium*, o conteúdo cultural codificado da imagem, o que ela comunica de modo previsível, do *punctum*, o detalhe que toca, que fura, que escapa ao código. A fotografia turística convencional é dominada pelo *studium*, pois reproduz o que já se sabe que deve ser visto, sem *punctum*, que perturbe o horizonte de expectativas. A possibilidade de uma fotografia turística com *punctum* seria a possibilidade de um turismo genuinamente transformador (Barthes, 1980/1984, p. 26-49).

O filósofo francês Jean Baudrillard (1981) leva essa ideia ao extremo. Para ele, vivemos na era dos simulacros. As cópias já não remetem a originais; elas próprias se tornaram a realidade. A Disneyland não é uma cópia da América; é uma América mais real do que a real, mais limpa, mais organizada, mais feliz. E, num certo sentido, a América real começa a se parecer com a Disneyland. Os shoppings, os condomínios fechados, os parques temáticos, tudo isso são simulacros que vão tomando conta do mundo. A teoria dos simulacros de Baudrillard (1981) articula-se com sua teoria da hiperrealidade, dado que, num estágio avançado do capitalismo de consumo, a distinção entre real e representação colapsa, e o simulacro precede e constitui o real. No turismo, isso significa que os destinos são construídos para corresponder às suas próprias imagens midiáticas, num círculo em que a representação não imita a realidade, mas a realidade imita a representação (Baudrillard, 1981/1994, p. 1-42).

O turista, nesse mundo, busca o autêntico mas só encontra simulacros. Quer ver a "Paris autêntica" e acaba na fila da Torre Eiffel. Quer ver a "cultura local" e acena num restaurante temático. Está preso num circuito de representações que não levam a lugar nenhum. É uma visão pessimista, talvez excessivamente pessimista. Afinal, as pessoas voltam das viagens transformadas. Trazem novas ideias, novas perspectivas, novas perguntas. Algo aconteceu ali, mesmo que não tenha sido o conhecimento factual, enciclopédico, que a escola valoriza.

O antropólogo Nelson Graburn (1989) propõe uma chave diferente. Para ele, o turismo é uma espécie de jornada sagrada (*sacred journey*), análoga à peregrinação religiosa. A gente sai do cotidiano (o mundo profano), viaja para lugares especiais (o mundo sagrado), vive experiências intensas que suspendem as regras normais (a liminaridade), e volta transformado para a vida co-

mum. O que se adquire na viagem não é tanto informação, mas renovação, sentido, encantamento.

A noção de liminaridade usada por Graburn (1989) é emprestada de Victor Turner (1969), que a desenvolveu para descrever os estados de transição nos rituais de passagem. O sujeito liminar está entre dois estados definidos, suspenso das hierarquias e obrigações normais, aberto à transformação. Turner (1969) mostrou que a peregrinação religiosa é uma das formas institucionalizadas de liminaridade nas sociedades complexas. Graburn (1989) argumenta que o turismo de lazer desempenha função estruturalmente equivalente nas sociedades secularizadas; ele é o ritual de transição que permite a renovação do sujeito ordinário (Turner, 1969, p. 94-113; Graburn, 1977, p. 17-31).

Essa ideia ressoa com a experiência de muita gente. A gente volta de viagem com uma sensação de ter vivido algo especial, de ter saído da rotina, de ter visto o mundo com outros olhos. Não é conhecimento proposicional, do tipo "saber que Paris tem 2 milhões de habitantes". É um conhecimento existencial, do tipo saber o que é sentir-se pequeno diante de uma montanha ou saber o que é perder-se numa cidade estrangeira. É um saber que não se traduz em palavras, mas que transforma quem somos.

Essa distinção entre conhecimento proposicional e conhecimento existencial retoma a diferença, já discutida na tradição pragmatista, entre saber-que (*knowing that*) e saber-como (*knowing how*). No contexto do turismo, o que está em jogo é um terceiro tipo, a saber, o conhecimento por familiaridade (*knowledge by acquaintance*), que William James (1890) distinguiu do conhecimento por descrição. Ter visto o azul do Mediterrâneo, tê-lo sentido, é diferente de saber proposicionalmente que é azul, e essa diferença é epistemicamente irredutível (James, 1890, p. 221-222). Isso não significa, porém, que o conhecimento factual seja irrelevante. Um turista que sabe algo sobre a história do lugar que visita, sobre a sua língua, sobre a sua cultura, vai ter uma experiência mais rica, mais profunda, mais significativa. O conhecimento não substitui a experiência; ele a alimenta, a contextualiza, a aprofunda. O problema é que o turismo de massa raramente proporciona esse conhecimento. As visitas são rápidas demais, as informações são superficiais demais, o contato com os locais é raro demais.

Há ainda a questão das relações de poder no conhecimento turístico. Quem fala sobre os lugares? Quem interpreta as culturas? Quem conta as histórias? Muitas vezes, são os guias, os livros, os sites, as operadoras, quase sempre pessoas de fora, com visões de fora. O conhecimento produzido sobre o lugar é, em grande medida, um conhecimento de estrangeiros para estrangeiros. Os locais raramente têm voz. Essa produção de conhecimento sobre o outro por agentes externos foi teorizada por Edward Said (1978) no conceito de orientalismo. O Ocidente produziu um *corpus* de saberes sobre o Oriente que serve menos para compreender as culturas orientais do que para legitimar o domínio sobre elas. No turismo, uma lógica análoga opera. O guia de viagem, o documentário, o blog etc., constroem uma imagem do destino que reflete as categorias e os desejos do turista, não a autopercepção das comunidades locais (Said, 1978, p. 3-28).

Pense num guia local, que cresceu naquela comunidade, que conhece as histórias que os livros não contam, que sabe onde as coisas realmente aconteceram. Muitas vezes, ele é visto apenas como um prestador de serviços, não como uma fonte de conhecimento. A sua palavra pesa menos do que a do historiador que nunca pisou ali. É uma forma de injustiça, sutil mas real. O turismo poderia ser diferente. Poderia haver mais espaço para as vozes locais, mais respeito pelos saberes tradicionais, mais troca genuína. Mas isso exige uma mudança de atitude, tanto dos turistas quanto das operadoras. Exige que a gente viaje não apenas para ver, mas para ouvir. Não apenas para consumir, mas para aprender. Não apenas para fotografar, mas para conversar.

4. Ética do Turismo

Agora vamos ao mais difícil. As questões éticas do turismo são as mais espinhosas, as que não têm resposta fácil, as que nos deixam desconfortáveis. Porque envolvem valores que colidem, interesses que se opõem, direitos que se contradizem. A ética do turismo não é uma subdisciplina marginal da ética aplicada; ela toca em questões fundamentais da filosofia moral e política, como os limites da liberdade individual quando ela afeta terceiros, a distribuição justa dos benefícios e custos de uma atividade econômica, o estatuto dos direitos cultu-

rais e do direito ao espaço urbano, e as responsabilidades que emergem de relações de poder assimétricas entre quem visita e quem é visitado.

Comecemos pela hospitalidade. O que significa receber alguém em casa? O filósofo Jacques Derrida (1997) pensou muito sobre isso. Ele diz que a hospitalidade tem duas faces. De um lado, há a lei incondicional da hospitalidade do dever de acolher o estrangeiro, o desconhecido, sem condições, sem perguntas, sem limites. É o ideal de abertura radical ao outro. Por outro lado, há as leis condicionais da hospitalidade; por exemplo, as regras concretas que estabelecem quem pode entrar, por quanto tempo, com que direitos, com que deveres. São as leis que protegem a casa, a comunidade, a vida em comum. A aporia da hospitalidade em Derrida (1997) não é facilmente solucionável, pois a hospitalidade incondicional é autodestrutiva, pois não distingue o hóspede do invasor, e a hospitalidade condicional é contraditória com o ideal de abertura radical. Derrida (1997) argumenta que toda prática concreta de hospitalidade opera na tensão entre os dois pólos, sem poder eliminá-la. Para o turismo, isso significa que não há fórmula que resolva o conflito entre abertura e proteção; qualquer política de gestão do turismo é uma negociação provisória nessa tensão insolúvel (Derrida; Dufourmantelle, 1997/2000, p. 25-55).

No turismo, essa tensão é vivida diariamente. As comunidades são pressionadas a acolher turistas incondicionalmente. O turismo é apresentado como salvação econômica, como oportunidade única, como caminho sem alternativa. Quem resiste é visto como atrasado, como hostil, como ingrato. Mas as comunidades também têm o direito de dizer não. De dizer "chega". De dizer que esta praia é nossa, esta rua é nossa, esta cidade é nossa. Quando os moradores de Barcelona vão às ruas gritar "turistas fora", eles estão exercendo esse direito. Estão dizendo que a hospitalidade tem limites, que a casa é deles, que eles decidem quem entra. O direito à cidade teorizado por Henri Lefebvre (1968) é relevante aqui, pois não se trata apenas do direito de acesso aos recursos urbanos, mas do direito de participar na produção do espaço urbano, de moldar a cidade conforme as necessidades e os valores de seus habitantes. O sobreturismo viola esse direito, ao impor uma produção do espaço urbano orientada para os visitantes, em detrimento dos residentes permanentes (Lefebvre, 1968, p. 167-175). O problema é que o turismo não é uma relação entre iguais. O turista tem dinheiro, tem passaporte, tem mobilidade. A comunidade, muitas

vezes, depende desse dinheiro, está presa àquele lugar, não tem para onde ir. A relação é assimétrica. O não da comunidade pode ter um custo alto.

Essa assimetria é estrutural. Ela reflete desigualdades globais de renda, mobilidade e poder político que o turismo não cria, mas amplifica. O turista do Norte Global que visita o Sul Global exerce uma forma de mobilidade privilegiada que a maioria dos habitantes do Sul jamais terá acesso. A filosofia política do turismo precisa confrontar essa assimetria em vez de pressupor que turistas e comunidades receptoras são agentes simétricos negociando em condições de igualdade (Bianchi, 2009, p. 484-508).

Conforme aludimos, o sobreturismo (*overtourism*) é a expressão mais visível dessa tensão. Os moradores perdem a sua cidade. Não conseguem mais morar onde nasceram, porque os preços dispararam. Não conseguem mais andar nas ruas, porque estão tomadas por visitantes. Não conseguem mais viver a sua vida, porque tudo é pensado para o turista. O sobreturismo é frequentemente tratado como um problema de gestão de fluxos, solucionável por instrumentos técnicos como tetos de visitantes ou taxas de entrada. Mas há uma dimensão mais profunda, na qual o sobreturismo é o resultado de um modelo econômico que externaliza os custos do turismo sobre as comunidades locais, enquanto privatiza os benefícios para os agentes econômicos do setor. A solução não é apenas técnica; é política, e requer uma redistribuição do poder de decisão sobre o uso do espaço (Milano et al., 2019, p. 1-8). O que fazer? Alguns defendem limites, como um número máximo de visitantes, ou taxas de entrada, ou restrições a novas licenças de alojamento local. Outros defendem uma redistribuição dos benefícios na qual o dinheiro do turismo chegue realmente às comunidades, que haja investimento em habitação acessível, em transporte público, em serviços para os moradores. Outros ainda defendem uma reconversão, a fim de, ao investir noutros setores, diversificar a economia e não depender tanto do turismo. Não há solução única. Cada lugar tem a sua história, a sua cultura, as suas possibilidades. O que não podemos é fingir que o problema não existe, ou achar que o mercado resolve tudo.

E ainda que o turismo gere riqueza, essa riqueza é mal distribuída. As grandes redes hoteleiras, as companhias aéreas, as plataformas de reserva ficam com a maior parte. As comunidades ficam com as migalhas e com os custos, como a poluição, degradação ambiental, inflação, descaracterização. A

economista Regina Scheyvens (2002) estudou esses vazamentos (*leakages*). Em muitos destinos, até 80% do dinheiro gasto pelos turistas sai da economia local. Vai para as multinacionais, para os paraísos fiscais, para os acionistas. O que fica é pouco. E o pouco ainda é mal distribuído, indo para quem já tem mais, para os donos dos grandes empreendimentos, para as elites locais. O fenômeno dos vazamentos econômicos é uma manifestação local de uma dinâmica global descrita pela teoria da dependência, a saber, nas relações entre economias centrais e periféricas, o excedente produzido na periferia é apropriado pelo centro através de mecanismos de troca desigual. No turismo, as economias receptoras, frequentemente no Sul Global, são integradas nas cadeias globais de valor em posição subordinada, retendo menos benefícios do que as empresas do Norte que controlam as plataformas de distribuição e as marcas (Scheyvens, 2002, p. 54-78).

O turismo de base comunitária (*community-based tourism*) tenta inverter essa lógica. A ideia é que a própria comunidade controle os empreendimentos, decida o desenvolvimento, retenha os benefícios. São os moradores que atuam como guias, que oferecem hospedagem, que preparam a comida, que contam as suas histórias. O turista não é um consumidor anônimo, mas um hóspede. É uma ideia bonita, mas difícil. Como competir com as grandes redes? Como aceder aos mercados internacionais? Como garantir qualidade sem padronizar? Como evitar conflitos internos sobre quem ganha o quê? Como evitar que a comunidade se torne ela própria um produto, uma atração exótica para turistas abastados?

Os obstáculos não são apenas internos à comunidade; há também conflitos sobre a distribuição de benefícios e sobre a dificuldade de padronização sem massificação. Competir nos mercados internacionais exige acesso a plataformas digitais de distribuição controladas por oligopólios (Booking, Airbnb, Expedia) cujas comissões reproduzem, em escala digital, os mesmos vazamentos que o turismo comunitário pretende evitar. A questão política central é, portanto, a regulação dessas plataformas, não apenas o apoio a iniciativas comunitárias isoladas.

A antropóloga Amalia Cabezas (2009) alerta para esse risco. Mesmo o turismo "solidário", responsável, comunitário pode reproduzir relações coloniais. A comunidade é pressionada a performar pobreza, a performar tradição, a perfor-

mar autenticidade, para agradar ao visitante. O turista quer ajudar, mas a sua ajuda pode reforçar a dependência, a subordinação e a desigualdade. O conceito de violência epistêmica de Gayatri Spivak (1988) é pertinente aqui. Ele nos diz que a ajuda externa, quando não mediada pela voz e pela agência das próprias comunidades, pode silenciar os sujeitos que pretende beneficiar, ao substituir seu autoentendimento por categorias impostas de fora. No turismo comunitário, isso se manifesta quando a narrativa sobre a comunidade é produzida para o consumo do turista, não pela comunidade para si mesma (Spivak, 1988, p. 271-313). Há uma crítica mais radical, vinda dos estudos decoloniais. Linda Tuhiwai Smith (1999), intelectual maori, mostra como o conhecimento ocidental sempre tratou os povos indígenas como objetos de estudo, não como sujeitos de conhecimento. No turismo, isso se repete. Os nativos são objetos do olhar, coisas a serem vistas, fotografadas, consumidas. Raramente são sujeitos que falam, que contam a sua própria história, que decidem como querem ser vistos.

Smith propõe uma metodologia de pesquisa indígena que parte do princípio de autodeterminação, na qual as comunidades indígenas devem controlar não apenas os recursos materiais de seus territórios, mas também a produção de conhecimento sobre elas. Transposto para o turismo, isso sugere que um turismo genuinamente descolonizado exige que as comunidades controlem não apenas os serviços que oferecem, mas a narrativa sobre quem são, o que implica a capacidade de recusar o olhar do turista, de não serem vistas quando não querem, de serem sujeitos ativos da relação (Tuhiwai Smith, 1999, p. 142-162). O que seria, então, um turismo descolonizado? Seria aquele em que as comunidades controlam a sua própria representação; em que decidem o que mostrar, o que esconder, como contar a sua história; em que negociam com os turistas de igual para igual; em que o encontro não é de consumo, mas de troca. É um ideal, claro. Difícil de alcançar. Mas talvez valha a pena tentar.

A noção de encontro intercultural simétrico é um ideal regulador, no sentido kantiano de uma ideia que orienta a prática sem jamais ser plenamente realizável nas condições históricas existentes. Sua função não é descrever o que o turismo é, mas orientar criticamente reformas e práticas concretas, i.e., um encontro é mais ou menos simétrico conforme se aproxima ou se afasta desse ideal. Isso permite avaliar práticas sem cair no relativismo nem no utopismo.

Por fim, a ética ambiental. O turismo é um dos grandes emissores de gases de efeito estufa. Aviões, cruzeiros, carros, hotéis, tudo isso consome energia, emite carbono, aquece o planeta. Um único voo de longa distância pode emitir tanto quanto um cidadão médio de um país pobre num ano inteiro. O setor do turismo é responsável por aproximadamente 8% das emissões globais de gases de efeito estufa quando considerado o ciclo completo, transporte, alojamento, atividades. A aviação é o pior deles, dado que um voo transatlântico de ida e volta emite em torno de 1,7 tonelada de CO_2 equivalente por passageiro, correspondendo, em ordem de grandeza, às emissões anuais per capita de vários países de renda baixa. A dimensão ética não é apenas individual; é estrutural, pois os sistemas de preço não internalizam os custos climáticos do turismo aéreo, gerando um subsídio implícito à mobilidade de longo curso (Lenzen *et al*., 2018, p. 522-527). O que fazer? Compensar as emissões? Plantar árvores, financiar projetos de energia limpa? É melhor do que nada, mas é insuficiente. E pode ser uma forma de aliviar a consciência sem resolver o problema. Repare que a solução não é fácil. Talvez viajar mais devagar (*slow travel*)? Ficar mais tempo em cada lugar, usar transportes menos poluentes, conhecer com mais profundidade. É uma ideia atraente, mas nem sempre possível. Nem todo mundo tem tempo para viajar devagar. Então viajar menos, simplesmente? Essa é a posição mais radical, e a mais impopular. Dizer que o direito ao turismo de longa distância não é um direito universal, que é um privilégio de poucos, e que esse privilégio tem um custo ambiental que não podemos pagar. É difícil de engolir, mas talvez seja a mais pura verdade.

Essa posição pode ser enquadrada na perspectiva da ética do clima, segundo a qual, em um cenário de emergência climática, o princípio de responsabilidade diferenciada atribui obrigações mais onerosas aos maiores emissores. Como o turismo aéreo é predominantemente praticado pelos segmentos de mais alta renda, a restrição ao turismo de longa distância tem uma dimensão de equidade, pois é uma política que onera os privilegiados em benefício dos mais vulneráveis às mudanças climáticas (Gössling; Cohen, 2014, p. 196-206).

5. Considerações Finais

O turismo é uma coisa estranha. Leva milhões de pessoas a saírem das suas casas para ver o que não conhecem, para viver experiências diferentes, para buscar qualquer coisa que não há no dia a dia. O turismo transforma lugares, transforma pessoas, transforma relações. Traz coisas boas e coisas más, quase sempre ao mesmo tempo. A estranheza do turismo é filosoficamente produtiva. Ela indica que o turismo não é uma prática entre outras, mas um fenômeno que coloca em questão categorias fundamentais como pertencimento, identidade, autenticidade, conhecimento e justiça. Uma filosofia do turismo não é, portanto, a aplicação de categorias preexistentes a um campo específico; é um terreno em que essas categorias são postas à prova e, eventualmente, refinadas. A metafísica do turismo nos mostrou que os lugares turísticos não são entidades fixas. São construídos, performados, interpretados, disputados. A autenticidade que os turistas buscam é uma coisa escorregadia, que muda de sentido conforme a perspectiva. Não há essência a ser preservada; há processos a serem compreendidos. A epistemologia nos mostrou que o conhecimento adquirido nas viagens é ambivalente. Pode ser superficial, de confirmação, de déjà vu. Mas pode ser também existencial, transformador, profundo. Depende de como viajamos, do que levamos, do que estamos dispostos a ver e a ouvir. A ética nos mostrou os dilemas de hospitalidade contra o direito à cidade, de desenvolvimento contra preservação, e do turista contra o morador. Não há respostas prontas. Há negociações, conflitos, escolhas. Diferentes lugares, diferentes culturas, diferentes momentos exigem diferentes respostas. O que serve para Veneza pode não servir para a Bahia. O que funcionou num ano pode não funcionar no outro. A tarefa de quem pensa o turismo, de quem planeja o turismo, de quem faz turismo, é estar atento a essas diferenças, a essas nuances, a esse balanceamento entre comprometimentos ontológicos, epistêmicos e éticos.

Esse pluralismo informado pressupõe padrões normativos, os ideais reguladores da hospitalidade justa, do encontro simétrico, da distribuição equitativa de benefícios e custos, e os aplica com sensibilidade contextual. O turismo é uma das formas mais comuns de encontro entre diferentes no mundo contemporâneo. Pode ser um encontro superficial, de consumo rápido, de olhar distraído. Mas pode ser também encontro genuíno, de troca, de aprendizado. Depende de

nós. Esse “depende de nós” merece ser precisado. Não se trata apenas de uma escolha individual de cada turista, mas de escolhas coletivas sobre as estruturas econômicas, jurídicas e políticas que organizam o turismo. As condições para um turismo melhor não são criadas apenas por turistas mais reflexivos, mas por políticas que redistribuam o poder e os benefícios, que protejam os direitos das comunidades receptoras, que internalizem os custos ambientais e que ampliem a voz dos sujeitos que hoje são apenas objetos do olhar turístico. A filosofia do turismo não termina na análise; ela aponta para uma política.

10

Filosofia da Engenharia e da Tecnologia

1. Por que uma Filosofia da Engenharia e da Tecnologia?

Você já parou para pensar no que acontece quando um engenheiro projeta uma ponte? Não estou falando dos cálculos, das fórmulas, dos materiais. Estou falando de algo mais fundamental. O que significa dizer que uma ponte é "segura"? Segura para quem? Segura em que condições? Segura por quanto tempo? Essas perguntas parecem técnicas, mas quando a gente começa a desdobrá-las, percebe que elas carregam questões filosóficas profundas. Pense num engenheiro que projeta um aplicativo. Ele decide como as informações serão mostradas, quais botões aparecem, quais dados são coletados, como são armazenados. Essas decisões parecem puramente técnicas, mas moldam o comportamento dos usuários, influenciam as suas escolhas, afetam as suas vidas. O engenheiro está, sem talvez se dar conta, fazendo escolhas éticas e políticas.

Essa duplicidade, decisões que se apresentam como técnicas mas carregam implicações normativas, é o problema central da filosofia da engenharia e da tecnologia. Ela não é uma preocupação recente. Já em meados do século XX, pensadores como Lewis Mumford (1934/2010) alertavam para o fato de que as escolhas técnicas incorporam visões de mundo e relações de poder que tendem a tornar-se invisíveis precisamente porque se apresentam como decisões de eficiência, não de valor (Mumford, 1934/2010, p. 212-234).

A tecnologia está em toda parte. Molda a maneira como trabalhamos, como nos comunicamos, como nos relacionamos, como pensamos. Mas raramente paramos para pensar no que ela é. O que é um artefato tecnológico? Uma ponte é diferente de uma pedra? Sim, porque foi projetada com uma intenção, para cumprir uma função. Mas onde termina a intenção e começa a função? Se alguém usa um martelo como peso de papel, ele ainda é um martelo?

E a relação entre tecnologia e valores? Um artefato pode ser neutro, ou ele carrega em si mesmo certos valores, certas visões de mundo? O historiador Winner (1980), num artigo clássico, perguntou se os artefatos têm política. (*Do Artifacts Have Politics?*). A pergunta parece estranha, mas Winner (1980) mostra que sim, artefatos podem incorporar relações de poder. As pontes baixas de Long Island, projetadas por Robert Moses, impediam a passagem de ônibus, dificultando o acesso de negros e pobres às praias. A política estava ali, no concreto, na altura da ponte.

A pergunta sobre a identidade funcional dos artefatos não é trivial; ela toca na questão filosófica sobre se a função é uma propriedade intrínseca do objeto ou uma relação entre o objeto, seus usuários e seus contextos de uso. A resposta tem consequências práticas, pois determina quando uma tecnologia pode ser considerada bem-sucedida, quando um design é responsável por seus efeitos e quando o uso desviante exonera o projetista.

Há também a questão do conhecimento. Que tipo de saber os engenheiros possuem? É o mesmo saber dos cientistas? Um físico sabe por que uma ponte cai; um engenheiro sabe como fazê-la não cair. São saberes diferentes, com métodos diferentes, com critérios de validação diferentes. A engenharia não é apenas "ciência aplicada", como às vezes se diz. Ela tem os seus próprios modos de conhecer, as suas próprias heurísticas, as suas próprias tradições.

Essa autonomia epistêmica da engenharia em relação à ciência é um tema central na filosofia da tecnologia contemporânea. Ela não implica que a engenharia prescinde do conhecimento científico, mas que o conhecimento da engenharia tem fontes, estruturas e critérios de validade próprios que não se reduzem aos de outras ciências amigas. Compreender essa autonomia é indispensável para avaliar corretamente a responsabilidade dos engenheiros, pois se o seu conhecimento é distinto, provavelmente as suas obrigações profissionais também o são.

E depois há a ética, a mais urgente de todas. Engenheiros projetam coisas que podem matar. Uma ponte mal calculada desaba. Um software mal escrito causa acidentes aéreos. Um algoritmo de IA reproduz preconceitos raciais. Quem é responsável? O engenheiro que escreveu a linha de código? O gestor que apressou o projeto? A empresa que priorizou o lucro? O sistema como um todo? A filósofa Helen Nissenbaum (1996) fala sobre o "problema das muitas

mãos", de que quando muitas pessoas contribuem para um resultado, fica difícil atribuir responsabilidade a alguém em particular.

A articulação das três dimensões, metafísica (o que são os artefatos), epistêmica (o que os engenheiros sabem) e ética (sobre o que eles são responsáveis), não é meramente expositiva. Elas se implicam mutuamente, dado que a resposta que se dá à questão sobre a natureza dos artefatos determina o escopo do que o engenheiro pode conhecer sobre seus efeitos, o que por sua vez delimita o alcance de sua responsabilidade. Um artefato cujos efeitos políticos são intrínsecos à sua estrutura gera obrigações de outro tipo das que gera um artefato cujos efeitos dependem inteiramente do uso.

Este capítulo é uma tentativa de pensar sobre tudo isso. Vamos começar pela metafísica, pelo que os artefatos tecnológicos são. Depois vamos para a epistemologia, para o que os engenheiros sabem. E vamos terminar na ética, com as perguntas mais difíceis sobre responsabilidade, justiça e controle. Não vou dar respostas prontas, pois o campo é demasiado complexo para isso. Mas talvez, ao final, as perguntas fiquem mais claras.

2. Metafísica da Engenharia e da Tecnologia

O que é um artefato tecnológico? A pergunta parece simples, mas não é. Um artefato não é apenas um objeto físico. Uma pedra é um objeto físico, mas não é um artefato. Um martelo é um artefato; foi feito com uma intenção, para cumprir uma função. A diferença está aí, na intencionalidade, na função.

Os filósofos da tecnologia Kroes & Meijers (2006) propõem uma teoria dual dos artefatos. Para eles, um artefato tem duas dimensões inseparáveis. De um lado, uma estrutura física, pois o martelo é feito de madeira e metal, tem um certo peso, um certo formato, uma certa resistência. De outro lado, uma função, dado que o martelo serve para martelar pregos, para quebrar coisas, para (quem sabe) servir de peso de papel. A estrutura física realiza a função através de processos causais. O peso do martelo, a dureza da cabeça de metal, o formato que permite segurar, tudo isso contribui para que ele cumpra a sua função.

A teoria dual de Kroes e Meijers (2006) insere-se num debate mais amplo sobre a redutibilidade do funcional ao físico. Ela assume que a descrição funcional de um artefato é irredutível à sua descrição física, i.e., conhecer toda a

física de um martelo não nos diz, por si só, que ele é um martelo, é preciso também conhecer as intenções do seu criador e as práticas sociais em que se insere. Essa irredutibilidade tem uma consequência importante para a engenharia, a de que o design de um artefato envolve sempre duas linguagens distintas, a da estrutura física e a da função, e a competência do engenheiro consiste precisamente em transitar entre elas (Kroes; Meijers, 2006, p. 5-27).

Mas a função não é uma propriedade física intrínseca. Um observador que nunca viu um martelo, vindo de uma cultura sem pregos, poderia ver ali apenas um pedaço de madeira com um peso na ponta. A função é atribuída, depende de um contexto de uso, de uma intenção de design, de uma prática social. O martelo é martelo porque foi feito para ser martelo, porque as pessoas o usam como martelo, porque existe uma prática de martelar pregos em que ele se insere.

A filósofa Beth Preston (1998) desenvolve essa ideia. Ela argumenta que a função de um artefato não é determinada apenas pela intenção do seu criador original, mas também pelos usos que a comunidade faz dele. Um artefato pode adquirir novas funções ao longo do tempo. O telefone foi inventado para comunicação de voz; hoje serve também para fotografar, navegar na internet, pagar contas, jogar. A função não é fixa; ela evolui com o uso.

A posição de Preston (1998) dialoga com a teoria da *affordance* de James J. Gibson (1979/2015), que argumenta que os objetos oferecem possibilidades de ação que são percebidas diretamente pelos agentes situados em determinados ambientes. Uma *affordance* não é uma propriedade puramente física do objeto nem uma projeção puramente subjetiva do agente, é uma relação entre as propriedades do objeto e as capacidades do agente num dado contexto. Transposta para a teoria dos artefatos, a noção de *affordance* sugere que a funcionalidade de um artefato é distribuída entre sua estrutura material, as capacidades dos usuários e os contextos de uso, o que explica tanto a estabilidade funcional (certos usos são favorecidos pela estrutura) quanto a plasticidade funcional (novos contextos revelam novas affordances) (Gibson, 1979/2015, p. 127-143). Isso tem implicações importantes. Se a função é determinada pelo uso, então os usuários têm um papel ativo na definição de o que os artefatos são. Eles não são meros consumidores passivos; são co-criadores de significa-

do. O design não termina na fábrica; continua na apropriação, na reinterpretação, na reinvenção cotidiana.

Artefatos complexos, como um avião ou um computador, têm uma organização hierárquica. São compostos de subsistemas, que são compostos de componentes, que são feitos de materiais. Cada nível tem a sua própria estrutura e a sua própria função. O motor do avião tem a função de gerar propulsão; as asas têm a função de gerar sustentação; os sistemas de controle têm a função de manter a estabilidade. O todo é mais do que a soma das partes, não no sentido místico, mas no sentido de que as relações entre as partes são tão importantes quanto as próprias partes.

Essa organização hierárquica dos artefatos complexos tem implicações epistêmicas e éticas que o texto desenvolve adiante, mas vale antecipar uma; a decomponibilidade hierárquica de um sistema não garante a previsibilidade do seu comportamento global. Herbert Simon (1969/1996) mostrou que sistemas hierárquicos são relativamente decomponíveis, o que permite que sejam projetados e analisados por partes, mas Charles Perrow (1984) mostrou que sistemas com alta complexidade interativa e acoplamento restrito geram comportamentos emergentes que escapam à previsão por partes (Simon, 1969/1996, p. 183-216; Perrow, 1984, p. 62-100).

Isso levanta a questão da identidade. Se um avião tem as peças substituídas uma a uma ao longo do tempo, ele continua sendo o mesmo avião? É o velho paradoxo do navio de Teseu, que já encontramos noutros capítulos. Para um engenheiro, o que importa é a continuidade funcional. Enquanto o avião continuar a voar, a cumprir a sua função, ele é o mesmo avião. A matéria pode mudar, a função persiste. Mas a função não é tudo. Um avião antigo, preservado num museu, já não voa. A sua função mudou; agora é testemunho histórico, objeto de contemplação. No entanto, continua a ser chamado de “o mesmo avião”. A identidade, aqui, parece depender mais da continuidade histórica, da cadeia de custódia, do que da função atual.

Esse ponto revela uma tensão na própria noção de artefato; ao mesmo tempo que a função é constitutiva do que o artefato é, a identidade histórica do artefato pode sobreviver à perda da função original. Essa tensão não é apenas acadêmica; ela tem consequências práticas para o direito de propriedade intelectual, para a conservação de patrimônio técnico e para a avaliação da respon-

sabilidade por artefatos que mudam de uso ao longo do tempo. Um software desenvolvido para fins civis que é readaptado para fins militares levanta questões análogas. É o mesmo artefato, com os mesmos criadores responsáveis, ou tornou-se algo distinto?

Mas passemos agora a uma questão mais espinhosa sobre se os artefatos incorporam valores. Eles são neutros ou carregam em si mesmos visões de mundo, relações de poder, escolhas políticas? Langdon Winner (1980), no seu artigo já mencionado, defende que sim. Ele distingue dois casos. No primeiro, há tecnologias que são intrinsecamente políticas, pois a sua própria natureza exige certos arranjos sociais. A energia nuclear, por exemplo, requer controle centralizado, hierarquias rígidas, segredo, devido aos riscos de acidentes e à possibilidade de proliferação de armas. A energia solar, por outro lado, é descentralizável, mais democrática, compatível com estruturas sociais igualitárias. Escolher uma ou outra não é apenas uma escolha técnica; é uma escolha sobre que tipo de sociedade queremos.

No segundo caso, há tecnologias que são contingentemente políticas, pois foram projetadas de determinada maneira para favorecer certos interesses. O exemplo de Winner (1980) são as pontes de Long Island, projetadas por Robert Moses propositadamente baixas para impedir a passagem de ônibus, que eram usados por negros e pobres para aceder às praias. A política estava ali, no concreto, na altura da ponte. Não é que pontes sejam intrinsecamente excludentes; esta ponte em particular foi projetada para excluir.

A distinção de Winner (1980) entre tecnologias intrinsecamente e contingentemente políticas é filosoficamente importante, mas foi criticada por ser excessivamente binária. Uma posição mais refinada, desenvolvida por Albert Borgmann (1984) em *Technology and the Character of Contemporary Life*, sustenta que as tecnologias modernas tendem a seguir um padrão, o que Borgmann (1984) chama de dispositivo, que sistematicamente oculta seus mecanismos internos, reduz o engajamento do usuário com o mundo e enfraquece as práticas culturais que davam sentido às atividades substituídas. Nessa leitura, a política da tecnologia não está apenas nos casos extremos de Winner (1980), mas no próprio modo de ser dos artefatos tecnológicos modernos (Borgmann, 1984, p. 40-68).

Críticos da posição de Winner (1980) argumentam que ele exagera. As tecnologias são flexíveis, podem ser reapropriadas, reconfiguradas. O automóvel, inicialmente visto como uma tecnologia que concentraria as pessoas nas cidades, acabou por permitir a expansão suburbana. A internet, projetada descentralizadamente, está a tornar-se cada vez mais centralizada nas mãos de poucas corporações. A política não está nos artefatos em si, mas nos usos que fazemos deles, nas instituições que os rodeiam, nas decisões coletivas que tomamos sobre eles.

Andrew Feenberg (2002), em *Transforming Technology*, propõe uma posição intermediária. Para ele, as tecnologias têm um código técnico, i.e., tendências incorporadas que favorecem certos usos e certos arranjos sociais. Mas esses códigos não são imutáveis; podem ser reformados através de contestação democrática e *redesign*. O automóvel, por exemplo, foi inicialmente projetado para velocidade e potência. Mas movimentos ambientalistas e de segurança conseguiram reformas como cintos de segurança, controle de emissões e eficiência energética. O código técnico foi alterado através de pressão social e regulação.

A noção de código técnico de Feenberg (2002) articula-se com a teoria crítica da tecnologia que ele desenvolve a partir de Herbert Marcuse e da tradição da Escola de Frankfurt. Para Feenberg (2002), a racionalidade técnica não é neutra; ela carrega, de forma latente, os interesses das classes dominantes que controlam o processo de desenvolvimento tecnológico. Mas diferentemente dos deterministas, Feenberg (2002) argumenta que essa racionalidade pode ser contestada e transformada por movimentos sociais que reivindicam uma racionalidade técnica alternativa, mais democrática, mais ecológica, mais atenta às necessidades dos trabalhadores e das comunidades afetadas. O design tecnológico é, portanto, um campo de luta política (Feenberg, 2002, p. 15-46).

O que isso significa para o engenheiro? Que ele não pode simplesmente alegar neutralidade. As suas escolhas de design têm consequências políticas, sociais e éticas. Ignorá-las não é ser neutro; é ser irresponsável.

Há ainda o debate entre determinismo tecnológico e construção social da tecnologia. O determinismo tecnológico, na sua versão mais forte, afirma que a tecnologia se desenvolve segundo uma lógica interna autônoma, e que uma vez desenvolvida, molda a sociedade de forma inevitável, por exemplo, dizendo-nos

coisas como “a internet tornou o mundo uma aldeia global”, como se a internet tivesse agência própria, como se as escolhas humanas não tivessem papel.

A abordagem SCOT (construção social da tecnologia) foi criticada por privilegiar os atores humanos em detrimento das resistências e contribuições dos próprios artefatos. Bruno Latour (1992) e a teoria ator-rede (ANT) propuseram uma simetria metodológica entre atores humanos e não humanos. Assim como os atores humanos moldam as tecnologias, as tecnologias também agem sobre os humanos, recrutando-os, resistindo a certos usos, permitindo outros. Numa formulação influente, Latour (1992) descreve os artefatos como “humanos duramente solidificados”, depósitos de decisões humanas passadas que continuam a agir no presente através de suas estruturas materiais. Esse enquadramento permite analisar simultaneamente as dimensões social e material da tecnologia sem reduzir uma à outra (Latour, 1992, p. 225-258).

A construção social da tecnologia foi desenvolvida por Bijker, Hughes e Pinch (1987), e ela argumenta que a tecnologia é moldada por fatores sociais. Diferentes grupos com diferentes interesses interpretam os artefactos de maneiras diferentes, e o design que se estabiliza é aquele que resulta dessas negociações. O exemplo clássico é a bicicleta, cujos designs iniciais variavam enormemente, e o que se tornou padrão não era inevitável; resultou de escolhas sociais.

Ambas as posições, levadas ao extremo, são insatisfatórias. O determinismo ignora a agência humana; o construtivismo social subestima as restrições materiais. Uma posição mais equilibrada reconhece a coprodução entre tecnologia e sociedade. A sociedade molda a tecnologia através de escolhas de investimento, regulação, design. Mas a tecnologia, uma vez implementada, molda a sociedade ao habilitar e restringir certas ações, criar dependências, influenciar percepções. A relação é dialética, não unidirecional. Essa coprodução dialética entre tecnologia e sociedade foi descrita por Sheila Jasanoff (2004) com o conceito de coprodução. A tecnologia e a ordem social são co-constituídas, ou seja, a produção de ordem social e a produção de artefatos tecnológicos são processos inseparáveis. Assim, a análise da tecnologia não pode ser separada da análise das instituições, das culturas e dos regimes de poder que a produzem (Jasanoff, 2004, p. 2-12).

3. Epistemologia da Engenharia

Passemos agora ao conhecimento. O que os engenheiros sabem que os cientistas não sabem? Ou, melhor, o saber da engenharia é diferente do saber da ciência? Uma visão comum, mas enganosa, é que a engenharia é "ciência aplicada". O cientista descobre as leis da natureza; o engenheiro aplica essas leis para construir coisas. Essa visão é enganosa por várias razões.

O primeiro a formular a inadequação do modelo "engenharia como ciência aplicada" de forma rigorosa foi o filósofo Mario Bunge (1966), que distinguiu entre ciência pura (orientada para a compreensão), ciência aplicada (orientada para a produção de conhecimento com aplicações potenciais) e tecnologia (orientada para a ação eficaz no mundo). Para Bunge (1966), a tecnologia não é apenas ciência aplicada, porque envolve um tipo de conhecimento, o conhecimento de como agir para atingir um resultado, que não é redutível ao conhecimento científico de como as coisas funcionam. A distinção é epistemologicamente importante, pois os critérios de sucesso tecnológico (o artefato funciona) diferem dos critérios de sucesso científico (a teoria é verdadeira) (Bunge, 1966, p. 329-349).

Muito do conhecimento dos engenheiros não deriva da ciência. Foi desenvolvido na prática, através de tentativa e erro, de experiência acumulada, de heurísticas transmitidas entre gerações. Vincenti (1990), no seu livro *What Engineers Know and How They Know It*, estudou a história da engenharia aeronáutica e identificou várias categorias de conhecimento técnico que são relativamente autônomas da ciência.

Há o conhecimento fundamental de design, i.e., os conceitos básicos que orientam a criação de artefactos. A ideia de que um avião voa porque as asas geram sustentação não é uma descoberta científica recente; é um conhecimento prático que foi sendo refinado ao longo de décadas de experimentação. Há o conhecimento quantitativo dos dados sobre o desempenho de materiais, sobre a resistência à fadiga, sobre o comportamento de estruturas sob carga. Esses dados são muitas vezes obtidos empiricamente, através de testes, e não deduzidos de teorias fundamentais. Há o conhecimento prático de saber como soldar, como inspecionar, como testar. É um saber que se adquire fazendo, que se transmite de mestre para aprendiz, que não está escrito em manuais. Há o co-

nhecimento de critérios, que é saber o que conta como uma ponte segura, o que conta como um design aceitável. Esses critérios são frequentemente normativos, envolvem julgamentos de valor, e não são claramente dedutíveis da ciência.

A taxonomia de Vincenti (1990) revela que o conhecimento de engenharia é estratificado em níveis com diferentes lógicas de produção e transmissão. Isso tem consequências para a questão da responsabilidade, pois, quando um sistema falha, a falha pode estar em qualquer um desses níveis, em um conceito de design equivocado, num dado empírico impreciso, num procedimento prático mal executado ou num critério normativo mal aplicado. Identificar o nível da falha é essencial para atribuir responsabilidade corretamente, mas a complexidade da estratificação torna essa identificação frequentemente difícil (Vincenti, 1990, p. 196-235).

A relação entre ciência e engenharia, portanto, não é de subordinação, mas de interação. Os engenheiros usam conhecimento científico, mas também produzem conhecimento que os cientistas usam. Os instrumentos que permitem a investigação científica são projetados por engenheiros. Os desafios práticos da engenharia frequentemente motivam a investigação científica. A termodinâmica foi desenvolvida em parte para compreender e melhorar as máquinas térmicas. A mecânica dos fluidos deve muito aos problemas da aviação.

Conforme falamos anteriormente, o filósofo Gilbert Ryle (1949) distinguiu dois tipos de conhecimento, o saber que (*knowing that*) e o saber como (*knowing how*). O saber que é proposicional, como o fato de que sei que Paris é a capital de França. O saber como é prático, como o fato de que sei como andar de bicicleta. A ciência é dominantemente um saber que, o saber proposicional; porém a engenharia é dominantemente um saber como. Um engenheiro pode não ser capaz de explicar todas as teorias por trás do que faz, mas sabe como fazer. Esse saber não é inferior; é simplesmente diferente.

Michael Polanyi (1966) falava em conhecimento tácito (*tacit knowledge*), aquele que não pode ser completamente articulado em palavras, que está incorporado nas práticas, nas habilidades, nos hábitos. Um engenheiro experiente reconhece um design problemático intuitivamente, sem ser capaz de explicar todas as razões. Esse conhecimento é real, é valioso, mas é difícil de transmitir em manuais.

A noção de conhecimento tácito de Polanyi (1966) é retomada por Harry Collins (2010) em sua teoria das "ondas do conhecimento" (*waves of knowledge*), que distingue vários tipos de tacitidade. Collins (2010) identifica tacitidade relacional (que pode ser tornada explícita por esforço ou mudança de contexto), tacitidade somática (incorporada no corpo) e tacitidade coletiva (distribuída na comunidade de prática e irredutível ao indivíduo). Na engenharia, as três formas estão presentes, pois o engenheiro aprende técnicas que pode eventualmente articular, desenvolve percepções perceptivas incorporadas que dificilmente verbaliza e participa de comunidades de prática cujos padrões tácitos sustentam seu julgamento profissional (Collins, 2010, p. 14-37).]

O design é a atividade central da engenharia. É o processo de criar algo novo para satisfazer determinados requisitos. Herbert Simon (1969), no seu livro *The Sciences of the Artificial*, caracterizou o design como uma busca num espaço de possibilidades. Temos um conjunto de requisitos funcionais e um conjunto de restrições. O espaço de design é o conjunto de todas as configurações possíveis de componentes e parâmetros que satisfazem as restrições. O problema é que esse espaço é, na maioria dos casos, astronomicamente grande, e não podemos explorá-lo exaustivamente.

A caracterização simoniana do design como busca em espaços de possibilidades foi refinada por estudos empíricos sobre o pensamento de design. Kees Dorst e Nigel Cross (2001) mostraram que o design não é apenas uma busca num espaço pré-definido, mas envolve a reformulação contínua do próprio problema, o que Dorst (2001) chama de co-evolução de problema e solução. O engenheiro não apenas busca soluções para problemas dados; ele redefine os problemas à medida que descobre restrições e possibilidades que não estavam explícitas no brief inicial. Esse caráter reflexivo do design distingue-o de outras formas de resolução de problemas e tem implicações para a formação de engenheiros (Dorst; Cross, 2001, p. 425-437).

Os engenheiros usam heurísticas para navegar nesse espaço. São regras práticas, atalhos, aproximações, que não garantem a solução ótima, mas tornam a busca viável. "Minimizar o número de componentes móveis" é uma heurística para aumentar a fiabilidade. "Usar materiais padrão quando possível" reduz custos. "Projetar para a fabricação" considera as capacidades dos pro-

cessos de produção. Essas heurísticas são aprendidas com a experiência, transmitidas entre gerações, e formam parte essencial do saber da engenharia.

A modelagem é outra ferramenta central. Modelos matemáticos, computacionais, físicos em escala permitem que os engenheiros prevejam o comportamento de sistemas, avaliem *trade-offs*, explorem alternativas, sem construir protótipos reais. Um engenheiro aeronáutico pode usar modelos computacionais de dinâmica de fluidos para testar centenas de variações de design de asa em horas, algo que seria impossível com protótipos físicos.

A epistemologia da modelagem em engenharia levanta questões filosóficas próprias. Os modelos não apenas representam sistemas existentes; eles também constituem o objeto do design ao tornar possíveis certas comparações e avaliações que definem o que conta como um design bem-sucedido. Isso tem implicações para a avaliação da responsabilidade, pois, quando um sistema falha de maneiras que os modelos não previram, a questão não é apenas se os modelos estavam "errados", mas se o espaço de possibilidades que eles constituíram era adequado ao problema real.

Mas todo modelo é uma simplificação. Omite detalhes considerados irrelevantes para o problema em questão. A questão é como sabemos que as simplificações são apropriadas. O estatístico George Box disse que "todos os modelos estão errados, mas alguns são úteis". A utilidade de um modelo depende do contexto. O mesmo sistema pode exigir modelos diferentes para propósitos diferentes. Um modelo de ponte adequado para análise estática pode ser inadequado para análise dinâmica sob carga de vento. O engenheiro precisa saber qual modelo usar em cada situação, e isso é, novamente, um conhecimento prático, não dedutível de regras gerais independentes de contexto.

Há ainda a questão da incerteza e do risco. Engenheiros projetam sistemas que devem funcionar em ambientes incertos. Eles não sabem exatamente que cargas uma ponte sofrerá ao longo da sua vida, não sabem exatamente as propriedades do solo onde a fundação será construída, não sabem exatamente como os utilizadores vão interagir com o produto. Como lidar com essa incerteza?

Tradicionalmente, os engenheiros usam fatores de segurança. Projetam a ponte para suportar muito mais do que a carga máxima esperada. Isso cria uma margem contra o desconhecido. Mas quão grande deve ser essa margem?

Muito pequena, e a ponte pode cair; muito grande, e a ponte fica cara demais. É uma escolha. E há diferentes tipos de incerteza. A incerteza aleatória é a variação inerente a processos naturais, e ela pode ser caracterizada probabilisticamente; a incerteza epistêmica é a ignorância sobre o estado do sistema, que pode ser reduzida com mais medições; por sua vez, a incerteza ontológica envolve eventos imprevisíveis a princípio, e ela não pode ser caracterizada probabilisticamente. O engenheiro precisa lidar com todos esses tipos.

O historiador Henry Petroski (1985), em *To Engineer Is Human*, mostra que muitas falhas catastróficas não resultam de erros de cálculo, mas de concepções erradas sobre como os sistemas funcionam. A ponte Tacoma Narrows colapsou não porque os materiais eram fracos, mas porque os engenheiros não compreenderam o fenômeno da excitação aerodinâmica. A incerteza aqui era epistêmica; não sabiam que não sabiam. E isso é mais difícil de gerir.

O caso da ponte Tacoma Narrows ilustra o que a literatura chama de ignorâncias de segunda ordem, aspectos do problema que o projetista não sabia que não sabia. Gerenciar incertezas epistêmicas é muito mais difícil do que gerenciar incertezas aleatórias; não há como calcular um fator de segurança para aquilo cuja existência se ignora. Isso sugere que a humildade epistêmica, a disposição de reconhecer os próprios limites de conhecimento, não é apenas uma virtude moral, mas uma necessidade técnica para o engenheiro que projeta sistemas sobre cujo funcionamento não tem controle completo (Petroski, 1985, p. 37-59).

Nancy Leveson (2011), em *Engineering a Safer World*, argumenta que as abordagens tradicionais de segurança, focadas na fiabilidade de componentes isolados, são inadequadas para sistemas complexos modernos. Acidentes emergem de interações inesperadas entre componentes, operadores humanos, software e ambiente. Para evitá-los, é preciso uma abordagem sistêmica, que considere o sistema como um todo, incluindo fatores organizacionais e humanos.

O acidente do ônibus espacial Challenger, por exemplo, não foi causado por uma falha aleatória de um componente, mas por uma combinação de problemas de design, pressões organizacionais para lançar, e decisões questionáveis de gestão. A responsabilidade estava distribuída por muitos agentes. É o que o cientista político Charles Perrow (1984) chama de acidentes normais (*normal*

accidents), pois em sistemas com alta complexidade interativa e acoplamento apertado, acidentes são inevitáveis, não importa quão bem projetados sejam.

Isso tem implicações epistemológicas profundas. Não podemos conhecer completamente sistemas complexos. Há limites para o que podemos prever e controlar. A humildade é, assim, uma virtude epistêmica essencial para engenheiros. A tese de Perrow (1984) sobre acidentes normais tem uma dimensão epistêmica que vai além da segurança técnica; ela implica que há um limite estrutural para o conhecimento que podemos ter sobre sistemas sociotécnicos complexos. Sistemas com alta complexidade interativa produzem comportamentos que emergem das interações entre componentes e que não podem ser deduzidos do comportamento de nenhum componente isolado. Isso não é apenas um desafio para a engenharia de segurança; é um limite epistêmico fundamental que tem implicações para a responsabilidade moral, dado que se o comportamento do sistema é estruturalmente imprevisível, a responsabilidade por seus efeitos não pode ser atribuída a nenhum projetista individual.

4. Ética da Engenharia e da Tecnologia

Chegamos às questões que realmente tiram o sono de engenheiros conscientes. Porque o trabalho deles não fica no papel; constrói pontes que podem desabar, escreve códigos que podem matar, projeta algoritmos que podem arruinar vidas. Uma ponte mal calculada não é apenas um erro teórico; é uma tragédia com nomes, rostos, famílias. Um software com bugs num sistema de controle de tráfego aéreo não é só um inconveniente; é uma questão de vida ou morte. Um algoritmo de concessão de crédito enviesado não é apenas uma imprecisão estatística; é uma família sendo empurrada para a exclusão financeira. Diante disso, a pergunta que se impõe é quem paga o preço quando a tecnologia falha.

Essa pergunta não é apenas moral; é também política e jurídica. Ela remete à distribuição do poder de impor riscos sobre outros sem o seu consentimento. A filosofia política identifica esse poder como uma das questões centrais da legitimidade do Estado e das instituições, a de saber quais riscos podem ser legitimamente impostos aos cidadãos e sob quais condições. A engenharia e a tecnologia tornaram-se um dos principais vetores de imposição de riscos nas

sociedades contemporâneas, o que torna a ética da engenharia inseparável da filosofia política da tecnologia.

Os códigos de ética profissional tentam dar uma resposta. Organizações como a National Society of Professional Engineers (NSPE) ou o Institute of Electrical and Electronics Engineers (IEEE) estabelecem princípios gerais, como o engenheiro ter de priorizar a segurança pública, atuar apenas dentro de suas áreas de competência, manter integridade e objetividade, evitar conflitos de interesse. São diretrizes nobres, sem dúvida. Mas quando você está lá, no chão da fábrica, pressionado por prazos, com a gerência ignorando suas preocupações, essas palavras bonitas não dizem o que fazer. Elas não resolvem o dilema.

A insuficiência dos códigos de ética para orientar a ação em situações de conflito foi diagnosticada por Michael Davis (1998) através do conceito de "problema de ética prática", i.e., situações em que princípios geralmente aceitos conflitam entre si ou são indeterminados quanto à sua aplicação. Davis argumenta que códigos de ética profissional funcionam melhor como instrumentos de coordenação, estabelecendo o que os membros da profissão podem esperar uns dos outros, do que como guias de ação em casos individuais. Para orientar a ação, é necessário um raciocínio ético contextualizado que vá além dos códigos, mobilizando tanto princípios gerais quanto julgamento situacional (Davis, 1998, p. 11-35).

Pense no caso clássico do *whistleblower*, o profissional que decide denunciar publicamente irregularidades na empresa onde trabalha. Um engenheiro descobre que um produto que sua equipe está desenvolvendo é inseguro. Os testes mostram falhas, os cálculos indicam risco, a intuição profissional grita que algo está errado. Mas a gerência, preocupada com o cronograma e com o lucro, decide ignorar os alertas e tocar o projeto adiante. O que esse engenheiro faz? Se denunciar internamente, pode ser ignorado ou silenciado. Se levar o caso para fora, para a imprensa ou para os órgãos reguladores, pode salvar vidas, mas também pode destruir sua carreira. A história de Roger Boisjoly, engenheiro da Morton Thiokol, é um lembrete brutal desse dilema. Ele alertou sobre os riscos dos anéis de vedação (os O-rings) do ônibus espacial Challenger em temperaturas baixas. Foi ignorado. O Challenger explodiu, matando sete astronautas. Boisjoly testemunhou perante a comissão de investigação, contou a

verdade, e por isso foi marginalizado na empresa, isolado pelos colegas, e nunca mais trabalhou como engenheiro. Fez a coisa certa? Moralmente, sim. Mas o preço que pagou foi imenso. O dilema do *whistleblower* não tem solução fácil; é um conflito entre lealdade à organização e responsabilidade com o público, entre a própria subsistência e o bem-estar de desconhecidos.

O caso Boisjoly expõe uma tensão estrutural nas profissões de engenharia entre a lealdade ao empregador (contratual e institucional) e a lealdade ao público (profissional e moral). A filosofia moral distingue entre obrigações *prima facie,* que valem em condições normais, mas podem ser suplantadas por obrigações mais urgentes, e obrigações absolutas. A denúncia pública de riscos graves é geralmente enquadrada como uma obrigação *prima facie* que pode tornar-se absoluta quando os canais internos se esgotam e os riscos são suficientemente sérios. Mas o preço individual da denúncia revela que o sistema moral que exige a denúncia sem proteger o denunciante é incoerente; não pode ao mesmo tempo exigir o sacrifício e tratar quem sacrifica como responsável pelas consequências.

Depois, há o problema do uso dual (*dual use*). Muitas tecnologias são ambíguas por natureza. Um sistema de criptografia pode proteger a privacidade de ativistas políticos em regimes autoritários, mas também pode esconder as comunicações de terroristas. Um drone pode ser usado para entregar medicamentos em áreas remotas ou para lançar mísseis em zonas de conflito. Um algoritmo de reconhecimento facial pode ajudar a encontrar crianças desaparecidas ou pode ser usado por governos para monitorar e oprimir minorias. Até onde vai a responsabilidade do engenheiro que desenvolveu a tecnologia? Se ele trabalha numa empresa que vende para o exército, ele é cúmplice dos usos militares? Se o software que ele escreveu é usado para vigilância em massa, ele tem culpa? Não há respostas prontas. Alguns argumentam que o engenheiro é responsável apenas pela qualidade técnica do seu trabalho, não pelos usos que outros farão dele. Outros rebatem que essa postura é uma forma de lavar as mãos, de fingir que a tecnologia é neutra quando ela nunca é.

O problema do uso dual é uma instância de um problema ético mais geral, a saber, a responsabilidade por consequências previsíveis de ações que são em si mesmas legítimas. A filosofia moral distingue entre o princípio da licitude indireta dos efeitos (uma ação com efeitos bons e maus pode ser permissível se

o efeito mau não for o meio para o efeito bom e se houver proporcionalidade entre eles) e a doutrina da contribuição culposa (quem contribui para um dano previsível com conhecimento das circunstâncias assume parte da responsabilidade pelo dano). No contexto do uso dual tecnológico, o segundo princípio parece mais pertinente, segundo o qual o grau de responsabilidade do engenheiro aumenta na medida em que os usos nocivos são mais previsíveis, a tecnologia é mais específica para esses usos e o engenheiro tem mais poder de influenciar o desenvolvimento ou a distribuição da tecnologia.

Há também a questão da incerteza. Projetar um sistema complexo é lidar com o desconhecido. O engenheiro não sabe exatamente como o solo vai se comportar daqui a cinquenta anos, não sabe que tipos de carga vão passar sobre a ponte, não sabe como os usuários vão interagir com o software. Os modelos são simplificações, os testes são limitados, o futuro é opaco. A gerência pressiona para lançar logo, os concorrentes estão na frente, o mercado não espera. Mais testes reduziriam a incerteza, mas custariam tempo e dinheiro. Que nível de certeza é suficiente? Não existe fórmula matemática que responda a isso. É um julgamento moral, uma aposta, uma decisão que pode custar vidas se for errada. O engenheiro que aprova o lançamento está assumindo um risco que não é só dele; é um risco que será imposto a usuários, a passageiros, a pedestres.

A questão de quanto teste é suficiente antes do lançamento é uma instância do problema filosófico da proporcionalidade entre risco e precaução. O princípio da precaução afirma que quando uma ação levanta ameaças de dano a seres humanos ou ao ambiente, medidas de precaução devem ser tomadas mesmo na ausência de certeza científica plena. Mas o princípio é controverso na engenharia porque pode levar à paralisia; qualquer lançamento envolve alguma incerteza, e exigir certeza completa antes de agir seria impedir a ação. A versão defensável do princípio de precaução na engenharia é relativa, a saber, o nível de certeza exigido deve ser proporcional ao nível de risco imposto e ao grau de reversibilidade dos danos possíveis.

O filósofo Davis (1998), em *Thinking Like an Engineer*, argumenta que a responsabilidade profissional dos engenheiros é diferente da moralidade comum. Não se trata apenas de ser uma boa pessoa; trata-se de ocupar uma posição de confiança, de ter um conhecimento especializado que outros não

têm, de ter o poder de causar danos em larga escala. Isso gera obrigações especiais. Mas Davis também tem clareza da realidade. Engenheiros não trabalham isolados; atuam em organizações complexas, com hierarquias, pressões e conflitos de interesse. Responsabilizar um indivíduo por falhas que são sistêmicas é injusto e ineficaz. É como culpar o soldado raso pela derrota na guerra.

O problema das muitas mãos levanta uma questão fundamental para a teoria moral, o de saber é possível atribuir responsabilidade a entidades coletivas, organizações, empresas, instituições, sem reduzir essa responsabilidade à soma das responsabilidades individuais dos membros. Philip Pettit e Christian List (2011) argumentam que sim. Grupos podem ser agentes morais genuínos quando têm crenças e desejos grupais que não coincidem com as crenças e desejos de nenhum membro individual, e quando têm procedimentos de tomada de decisão que produzem ações atribuíveis ao grupo como tal. Isso sugere que a NASA, como instituição, poderia ser moralmente responsável pelo desastre do Challenger de formas que nenhum indivíduo isolado poderia ser (List; Pettit, 2011, p. 23-54).

O cientista político Dennis Thompson (1980) deu um nome a esse fenômeno, o problema das muitas mãos. Em organizações complexas, com responsabilidades fragmentadas e distribuídas, resultados errados podem acontecer sem que nenhuma pessoa específica seja claramente culpada. Cada um fez sua pequena parte, seguiu os procedimentos, tomou decisões razoáveis no seu nível. A falha emerge da interação entre essas partes, não da ação de um indivíduo. No desastre do Challenger, quem é o responsável? Os engenheiros que projetaram o O-ring? Os que alertaram mas não conseguiram impedir o lançamento? Os gerentes que aprovaram mesmo com as advertências? A NASA como instituição, com sua cultura organizacional que silenciava dissidências? Não há um único culpado; há uma teia de decisões, omissões, pressões e falhas de comunicação.

Helen Nissenbaum (1996) aprofunda essa análise, identificando barreiras específicas à responsabilização em sistemas computacionais. Além da própria difusão da responsabilidade, o problema das muitas mãos, que já mencionamos, temos também a inevitabilidade de bugs, pois um software complexo sempre tem falhas, e esperar que seja perfeito é irrealista. A opacidade dos algoritmos, especialmente com a ascensão do aprendizado de máquina, onde nem os

próprios criadores entendem completamente por que o sistema tomou determinada decisão, também é um ponto problemático, tal como a propriedade privada dos sistemas, que levam a códigos fechados e segredos comerciais que impedem auditorias independentes. Essas barreiras não eliminam a responsabilidade, mas a tornam difusa, difícil de rastrear, fácil de negar. As soluções propostas vão desde modelos de responsabilidade coletiva ou organizacional (a empresa como um todo responde) até exigências de transparência e auditabilidade (sistemas precisam ser inspecionáveis), passando pela participação das partes interessadas e pela regulação proativa (em vez de esperar o desastre para agir).

As quatro barreiras identificadas por Nissenbaum (1996) podem ser reordenadas segundo distintas dimensões da responsabilidade, o que permite explicitar o tipo de imputação comprometido em cada caso. A difusão incide sobre a responsabilidade causal, na medida em que torna indeterminável a atribuição do dano a um agente específico. A inevitabilidade dos bugs compromete a responsabilidade por negligência, pois enfraquece a possibilidade de sustentar que o resultado poderia ter sido evitado. A opacidade afeta a responsabilidade por intenção, uma vez que obscurece a relação entre a ação executada e o saber do agente acerca de seus próprios procedimentos. A propriedade privada, por sua vez, limita a responsabilidade como prestação de contas, ao restringir as condições de verificação externa do comportamento do agente.

Dessa diferenciação decorre que as respostas normativas não são homogêneas. Medidas como transparência e auditabilidade incidem diretamente sobre os problemas de opacidade e de propriedade privada, ao ampliar as condições de inspeção e controle. Elas permanecem, contudo, insuficientes diante da difusão da agência e da inevitabilidade de falhas em sistemas complexos, nas quais a dificuldade não reside apenas no acesso à informação, mas na própria estrutura distribuída da ação e na contingência técnica do funcionamento (Nissenbaum, 1996, p. 26-52).

Há ainda as questões de justiça distributiva, que raramente entram nas discussões técnicas mas são fundamentais. Toda tecnologia gera benefícios e impõe riscos. A pergunta é quem fica com os benefícios e quem arca com os riscos. A história mostra um padrão recorrente. As comunidades ricas e poderosas ficam com os ganhos; as comunidades pobres e marginalizadas ficam com

os custos. Incineradores de lixo, aterros sanitários, indústrias poluentes, todos tendem a ser instalados em áreas de baixa renda, frequentemente habitadas por minorias étnicas. Nos Estados Unidos, esse fenômeno tem um nome, a saber, injustiça ambiental (*environmental injustice*). Não é acaso; é resultado de decisões de planejamento, de zoneamento, de licenciamento, que sistematicamente empurram os riscos para quem tem menos poder de reclamar. O racismo ambiental é uma forma de racismo estrutural, mediado por decisões técnicas que se apresentam como neutras.

Kristin Shrader-Frechette (1991), em *Risk and Rationality*, defende que decisões sobre riscos tecnológicos não podem ser deixadas apenas nas mãos de especialistas. Técnicos podem avaliar magnitudes de risco, calcular probabilidades, modelar cenários. Mas decidir se um risco é aceitável não é uma questão técnica; é uma questão de valor. Envolve pesar o que uma comunidade está disposta a sacrificar em troca de desenvolvimento, e perguntar quem está sendo protegido e quem está sendo exposto. Por isso, Shrader-Frechette (1991) argumenta que as decisões devem ser democráticas.

O argumento de Shrader-Frechette (1991) articula-se com a distinção epistêmica entre perícia e julgamento de valor. A perícia técnica autoriza pronunciamentos sobre a magnitude dos riscos, mas não sobre a sua aceitabilidade. A segunda questão é normativa e exige a participação das comunidades afetadas. Mas Shrader-Frechette (1991) argumenta que a própria definição de risco, o que conta como dano, que escala temporal é relevante, que populações devem ser consideradas, já envolve julgamentos de valor que os especialistas frequentemente tomam sem reconhecer seu caráter normativo. A democratização das decisões de risco deve incluir, portanto, a democratização da própria definição do que é um risco (Shrader-Frechette, 1991, p. 183-215).

O problema, claro, é que essa exigência esbarra em dificuldades práticas bem conhecidas. Os cidadãos comuns não têm o treinamento técnico dos especialistas. As empresas e os governos, por sua vez, controlam informações que não são facilmente acessíveis. Assim, os processos participativos tendem a ser lentos, custosos, e podem ser capturados por grupos organizados que não representam a comunidade como um todo. Não há uma solução simples para esse dilema. Alguns apostam na deliberação pública informada como caminho, ao criar espaços onde especialistas apresentem informações de forma acessí-

vel, onde cidadãos possam fazer perguntas e onde diferentes perspectivas sejam efetivamente ouvidas. Conselhos de cidadãos, audiências públicas e conferências de consenso são tentativas de colocar essa ideia em prática. Mas mesmo essas iniciativas enfrentam limites sobre quem participa, quem fala em nome de quem, como garantir que a deliberação não seja apenas mais uma formalidade.

Outro problema de justiça que ganhou imensa visibilidade nos últimos anos é a exclusão digital (*digital divide*). O acesso à internet, aos computadores e aos smartphones não é universal. No mundo, bilhões de pessoas ainda estão desconectadas. Nos países ricos, a divisão se dá por renda, por educação, por localização geográfica. Comunidades pobres, rurais, periféricas têm acesso pior, mais caro, mais lento. As consequências são profundas. Oportunidades educacionais, empregos, serviços públicos, participação política; tudo hoje depende crescentemente do acesso digital. A pandemia de COVID-19 escancarou isso. Crianças sem internet em casa simplesmente pararam de estudar durante o ensino remoto. Jovens sem computador ficaram de fora do mercado de trabalho. Idosos sem familiaridade com aplicativos não conseguiram nem agendar suas vacinas. A exclusão digital não é apenas um problema técnico que o mercado resolveria se fosse eficiente; é uma questão de justiça distributiva que exige políticas públicas, investimento em infraestrutura, programas de inclusão, subsídios. Mas também aqui não há consenso sobre o desenho dessas políticas, nem sobre quem deve pagar a conta.

A exclusão digital pode ser analisada a partir da teoria das capacidades de Amartya Sen (1999) e Martha Nussbaum (2011). Na abordagem das capacidades, o que importa para a justiça não é apenas a distribuição de recursos, mas a conversão de recursos em capacidades reais de ação. O acesso à internet é um recurso que, para ser convertido em capacidade, requer habilidades digitais, dispositivos adequados, infraestrutura de rede e conteúdo relevante. Comunidades que carecem de qualquer um desses elementos não se beneficiam do acesso formal, mesmo quando ele existe. Isso implica que políticas de inclusão digital não podem limitar-se a prover acesso técnico; devem abordar todo o conjunto de fatores que convertem o acesso em capacidade real (Sen, 1999, p. 70-86; Nussbaum, 2011, p. 17-45).

Há ainda uma dimensão mais profunda, que diz respeito à própria natureza da vida social. Algumas tecnologias aumentam a autonomia humana. Ferramentas que expandem nossas capacidades, que nos permitem fazer coisas que não podíamos, que nos conectam com outras pessoas, que nos dão mais controle sobre nossas vidas. Outras tecnologias fazem o oposto, pois reduzem nossa autonomia. Sistemas de vigilância que limitam nossa privacidade, automação que tira nossos empregos, algoritmos que decidem por nós o que vemos, o que compramos, o que pensamos. Langdon Winner (1980), no seu trabalho clássico, distinguia tecnologias autoritárias, que concentram poder e exigem controle centralizado, de tecnologias democráticas, que distribuem poder e facilitam a participação. Escolher que tecnologias desenvolver, como desenvolvê-las, como implementá-las, isso é escolher que tipo de sociedade queremos. Não é uma decisão puramente técnica; é uma decisão política, moral, existencial.

Um tema que atravessa a filosofia crítica da tecnologia, desde Marx (1844/2004), é o da alienação. Nas fábricas industriais do século XIX, Marx (1844/2004) via trabalhadores reduzidos a apêndices das máquinas, separados do produto do seu trabalho, dos processos de produção, da sua própria criatividade. Essa crítica, longe de estar datada, parece mais atual do que nunca. Trabalhadores de *call center* seguem *scripts* algorítmicos que determinam cada palavra que devem dizer. Motoristas de aplicativo são monitorados, avaliados, dirigidos por sistemas que controlam cada curva, cada parada, cada minuto. Funcionários de armazéns têm seus movimentos otimizados por algoritmos que calculam a rota mais eficiente para separar mercadorias, ignorando o cansaço, a individualidade e a dignidade. Há uma perda real de autonomia, de criatividade e de sentido no trabalho. A tecnologia, que poderia libertar, muitas vezes aprisiona.

Andrew Feenberg (2002), em *Transforming Technology*, argumenta que essa não é uma fatalidade. Tecnologias podem ser redesenhadas, reformadas, reorientadas para aumentar a autonomia em vez de reduzi-la. O software de código aberto é um exemplo. Qualquer um pode inspecionar o código, modificá-lo, adaptá-lo às suas necessidades, compartilhar suas modificações. O usuário não é um consumidor passivo, mas um participante ativo. O software proprietário, ao contrário, trata o usuário como alguém que apenas consome o que lhe é dado, sem poder de intervenção. A diferença não é apenas técnica; é política. O

design tecnológico é um campo de disputa, onde movimentos sociais podem pressionar por tecnologias mais democráticas, mais participativas, mais respeitosas da autonomia humana. Engenheiros, nesse contexto, não são meros executores de ordens; são agentes que podem escolher projetar para a autonomia ou para o controle.

A noção marxiana de alienação refere-se primariamente à separação do trabalhador em relação ao produto do seu trabalho, ao processo de produção, a outros seres humanos e à sua própria natureza genérica como ser criativo (Marx, 1844/2004, p. 111-127). Na tecnologia contemporânea, essas quatro dimensões da alienação reaparecem em formas novas. A separação do produto é agravada pela invisibilidade do trabalho digital; a separação do processo é agravada pela mediação algorítmica, que remove do trabalhador o controle sobre o próprio processo; a separação dos outros é agravada pela individualização das plataformas; e a separação da natureza genérica é agravada pela redução do trabalho à execução de procedimentos pré-definidos. Feenberg (2002, p.152-180) argumenta que a tecnologia de código aberto e outras formas de produção participativa representam tentativas de superar a alienação tecnológica através do redesign democrático dos artefatos.

Uma categoria especial de tecnologia merece atenção à parte, a saber, os sistemas autônomos, que incluem inteligência artificial, robótica e veículos autônomos. Esses sistemas tomam decisões sem intervenção humana direta. E aprendem, se adaptam, mudam seu comportamento com base em dados. Isso levanta questões éticas de uma ordem completamente nova. Se um carro autônomo atropela um pedestre, quem é o responsável moral e legal? O fabricante, que projetou o sistema? O programador, que escreveu o código? O proprietário, que comprou o carro? O próprio sistema, se é que podemos atribuir-lhe algum tipo de agência? Não há resposta clara, e os tribunais ao redor do mundo estão começando a enfrentar essas questões sem ter precedentes claros. O problema das muitas mãos, que já discutimos, se agrava. O sistema aprendeu com milhões de exemplos; seu comportamento não foi explicitamente programado por ninguém. O engenheiro que projetou a arquitetura da rede neural não pode prever todas as decisões que ela tomará. O cientista de dados que selecionou os dados de treinamento não pode garantir que eles não contenham vieses ocultos. O gestor que aprovou o projeto não entende os detalhes técnicos.

Quando o sistema falha, é quase impossível apontar um único responsável. Andreas Matthias (2004) chamou isso de lacuna de responsabilidade (*responsibility gap*). Há um vazio onde a responsabilidade deveria estar, mas não está. O comportamento emerge do treinamento, não do design deliberado. Ninguém previu exatamente aquela decisão. Ninguém pode ser culpado por ela no sentido tradicional.

A lacuna de responsabilidade identificada por Matthias (2004) pode ser esclarecida pela distinção entre imputação pelo dano e imputação por deveres de cuidado. A primeira vincula a responsabilidade à produção efetiva de um prejuízo, ao passo que a segunda a fundamenta na observância ou violação de procedimentos adequados de precaução. Matthias (2004) sustenta que sistemas de aprendizado de máquina instauram uma lacuna na imputação pelo dano, uma vez que o comportamento do sistema não é integralmente previsível pelos projetistas. Essa indeterminação, contudo, não suprime a imputação por deveres de cuidado. Fabricantes, desenvolvedores e operadores permanecem obrigados a assegurar padrões adequados de teste, validação, monitoramento e eventual desativação. A dificuldade de atribuição individual do dano pode ser parcialmente compensada por regimes de responsabilidade objetiva, independentes de culpa, que asseguram a reparação das vítimas mesmo na ausência de imputação subjetiva determinada (Matthias, 2004, p. 175-183).

Diante disso, alguns autores, especialmente no direito, propõem uma saída radical, que seria tratar os próprios sistemas autônomos como agentes morais, pelo menos num sentido funcional. Se um sistema é capaz de avaliar cursos de ação, tomar decisões com base em critérios, e agir no mundo, talvez deva ser responsabilizado por suas ações. Isso implicaria criar uma nova categoria jurídica, algo como “pessoas eletrônicas”, com personalidade limitada e capacidade de responder civilmente por danos. A ideia é intrigante, mas enfrenta objeções formidáveis. Que tipo de responsabilidade um sistema de IA poderia ter? Multá-lo não faz sentido, já que dinheiro não significa nada para ele. Desligá-lo é punição ou solução técnica? E, principalmente, não seria isso apenas uma maneira de desviar a atenção dos verdadeiros responsáveis? Se uma empresa lucra com um sistema autônomo que causa danos, não deveria ser a empresa a responder, em vez de transferir a responsabilidade para o sistema?

A posição mais defensável, pelo menos por enquanto, parece ser que a responsabilidade moral por sistemas autônomos recai sobre os humanos que os projetam, desenvolvem, implementam e operam. Mas à medida que os sistemas se tornam mais autônomos e suas decisões mais imprevisíveis, esse modelo começa a mostrar suas limitações. Talvez tenhamos que pensar em termos de responsabilidade corporativa ou institucional, em vez de individual. Talvez tenhamos que criar mecanismos de compensação automática, seguros obrigatórios, fundos de reparação. O importante é não fingir que o problema não existe, nem esperar que a tecnologia se autorregule.

Nesse sentido, não podemos esquecer a questão do viés algorítmico, que já discutimos em capítulos anteriores, mas merece ser retomada aqui do ponto de vista da engenharia. Sistemas de IA treinados em dados históricos aprendem os padrões desses dados. Se os dados refletem uma discriminação passada (mulheres contratadas em menor número, negros presos desproporcionalmente), o algoritmo vai reproduzir e amplificar esses vieses. Um sistema de recrutamento pode "aprender" que homens são melhores para certas posições, simplesmente porque no passado a empresa contratou mais homens. Um sistema de reconhecimento facial pode ter taxas de erro muito mais altas para pessoas negras, porque foi treinado majoritariamente com fotos de pessoas brancas. O problema não é que os algoritmos sejam racistas ou sexistas no sentido humano; eles não têm intenções. O problema é que eles tornam a discriminação automática, invisível, difícil de contestar. Engenheiros de software têm a responsabilidade de identificar e mitigar esses vieses, mas muitas vezes não têm controle sobre os dados de treinamento, ou sobre as decisões de negócio que moldam o produto.

Isso nos leva à questão da transparência e da explicabilidade. Modelos de deep learning são, em grande parte, caixas-pretas. Recebem dados, produzem resultados, mas é quase impossível explicar como chegaram àquela decisão. Um radiologista que usa um sistema de IA para diagnosticar câncer pode confiar nele, mas não pode explicar ao paciente por que o sistema deu aquele resultado. Um juiz que usa um algoritmo para recomendar sentenças não pode fundamentar sua decisão no funcionamento interno do modelo. As dificuldades são múltiplas. A confiança cega não é justificável; se não entendemos como o sistema funciona, não sabemos quando ele pode falhar. Além disso, temos exigências legais de explicabilidade; o Regulamento Geral de Proteção de Dados

(GDPR) europeu, por exemplo, estabelece um "direito à explicação" para decisões automatizadas. Sem explicabilidade é impossível contestar decisões injustas. Se o sistema negou um empréstimo, e não sabemos por quê, como recorrer?

Engenheiros enfrentam aqui um dilema sem solução óbvia. Modelos mais complexos, como redes neurais profundas, costumam ser mais precisos, mas são opacos. Modelos mais simples, como regressões logísticas ou árvores de decisão, são mais interpretáveis, mas podem ser menos precisos. Cynthia Rudin (2019) defende que, para aplicações de alto risco, devemos priorizar a interpretabilidade. Melhor um modelo um pouco menos preciso, mas que possamos entender e auditar, do que uma caixa-preta que não podemos contestar. Outros argumentam que, em certos domínios, a precisão salva vidas, e que devemos aceitar a opacidade em troca de melhores resultados. O debate está longe de resolvido, e provavelmente não há uma resposta que sirva para todos os casos.

O debate entre precisão e interpretabilidade em modelos de IA pode ser reformulado, no plano ético, como um conflito entre valores distintos. De um lado, a eficácia, entendida como maximização de acertos. De outro, a justiça procedimental, que exige processos decisórios compreensíveis e passíveis de contestação. Rudin (2019) sustenta que essa oposição é frequentemente superestimada. Em diversos contextos de alto risco, modelos interpretáveis bem projetados alcançam níveis de precisão próximos aos de sistemas opacos, sobretudo quando o problema apresenta estrutura definida. Nesses casos, a escolha não implica perda relevante de desempenho. Há situações, contudo, em que a tensão persiste. Nelas, a filosofia moral tende a atribuir primazia à justiça procedimental quando estão em jogo direitos fundamentais. Esse deslocamento normativo justifica a preferência por modelos interpretáveis em decisões que incidem sobre crédito, emprego, saúde e justiça criminal (Rudin, 2019, p. 206-215).

5. Considerações Finais

A engenharia e a tecnologia moldam o mundo em que vivemos de maneiras tão profundas que mal percebemos. As pontes que atravessamos, os edifícios que habitamos, os dispositivos com que nos comunicamos, os algoritmos que decidem o que vemos, o que compramos, o que pensamos; tudo isso é produto

do trabalho de engenheiros. Mas esse trabalho não é neutro. Cada projeto envolve escolhas, e cada escolha carrega valores, implica consequências, distribui riscos e benefícios.

A persistência da ilusão de neutralidade da engenharia e da tecnologia não é meramente um erro cognitivo; é um fenômeno social com funções específicas. Ela isenta os projetistas de responsabilidade por consequências previsíveis, legitima decisões técnicas que de outro modo seriam reconhecidas como políticas e impede a participação democrática em decisões que afetam a vida de comunidades inteiras. A filosofia da engenharia e da tecnologia tem, portanto, uma função crítica, a de tornar visível o que a ilusão de neutralidade oculta.

Na metafísica, vimos que os artefatos tecnológicos têm uma natureza dupla. São, ao mesmo tempo, estruturas físicas, com massa, resistência, propriedades materiais, e entidades funcionais, com propósitos, usos, significados. A função não está na matéria, mas na relação entre a matéria e as intenções humanas, as práticas sociais, os contextos de uso. Um mesmo objeto físico pode ter funções completamente diferentes em diferentes culturas, épocas, situações. E os artefatos podem incorporar valores, cristalizar relações de poder e também materializar visões de mundo. A tecnologia não é apenas ferramenta; é política.

O corolário metodológico dessa conclusão metafísica consiste em ampliar o escopo da análise dos artefatos tecnológicos. Não basta descrever propriedades físicas ou funções técnicas. É necessário examinar as relações de poder que orientam o design, as práticas sociais que estabilizam os usos e as narrativas culturais que legitimam ou contestam a existência do artefato. Esse deslocamento não introduz um nível "social" externo a uma base "técnica" mais fundamental. Trata-se de uma exigência interna à própria ontologia dos artefatos. Tais entidades não se deixam reduzir a um único registro. Elas se constituem simultaneamente como configurações materiais e como formações culturais.

Na epistemologia, vimos que o saber da engenharia é distinto do saber da ciência. Não se trata de aplicar teorias pré-existentes, mas de construir conhecimento na prática, através de heurísticas, modelos, experimentação, tentativa e erro. O engenheiro sabe como fazer, mesmo quando não sabe explicar todas as razões. Esse conhecimento tácito é tão real quanto o conhecimento explícito, mas é mais difícil de transmitir, de documentar, de auditar. E há limites para o que podemos saber. Os sistemas complexos são imprevisíveis, os modelos são

sempre simplificações, o futuro é incerto. A humildade epistêmica é uma virtude essencial para quem projeta coisas que podem matar se falharem.

Essa humildade epistêmica não se reduz a uma virtude individual do engenheiro responsável. Ela deve ser incorporada às estruturas institucionais das organizações que desenvolvem e operam sistemas tecnológicos. Isso envolve a consolidação de culturas organizacionais que valorizem a identificação e a comunicação de incertezas, em vez de sua supressão. Implica também a adoção de processos de design que incluam, de modo explícito, margens para incertezas ontológicas, para além das aleatórias. Requer ainda a implementação de mecanismos de monitoramento contínuo capazes de aprender com falhas imprevistas, sem reduzir tais ocorrências à culpabilização individual diante de eventos estruturalmente inevitáveis. Nessa perspectiva, a segurança sistêmica, conforme formulada por Leveson (2011), depende da institucionalização da humildade epistêmica.

A humildade epistêmica não pode ser compreendida apenas como uma disposição individual do engenheiro. Ela precisa ser incorporada ao nível institucional das organizações que desenvolvem e operam sistemas tecnológicos. Essa incorporação se manifesta, em primeiro lugar, na formação de culturas organizacionais que incentivam a identificação e a comunicação de incertezas, em vez de sua ocultação. Em seguida, exige processos de design que reconheçam explicitamente a presença de incertezas ontológicas, sem reduzi-las às variações aleatórias. Além disso, requer a implementação de formas contínuas de monitoramento que permitam aprender com falhas imprevistas, evitando a tendência de atribuí-las exclusivamente a indivíduos quando decorrem de condições estruturais. Nesse quadro, a noção de segurança sistêmica em Leveson (2011) depende diretamente da institucionalização dessa forma de humildade epistêmica.

Esses dilemas não constituem exceções que surgem apenas quando a engenharia falha. Eles pertencem à própria estrutura de práticas técnicas que operam em larga escala e sob condições de incerteza. Por essa razão, a ética da engenharia não pode limitar-se a uma função corretiva. Ela deve assumir um caráter prospectivo e acompanhar o processo desde a sua formulação inicial.

Essa orientação implica inserir critérios éticos no próprio desenvolvimento técnico, e não apenas avaliá-lo ao final. A literatura descreve esse movimento

por expressões como *ethics by design* e *values in design*. Em termos mais precisos, trata-se de incorporar valores normativos como critérios internos de projeto, que orientam escolhas, restrições e prioridades ao longo do desenvolvimento. A ética, nesse quadro, não atua como instância externa de validação. Ela participa da própria configuração do artefato, desde as primeiras decisões de concepção.

O pluralismo que emerge dessas reflexões é uma constatação prática. Não há uma resposta única para todos os problemas, um código de ética que resolva todos os dilemas, uma fórmula matemática que calcule a decisão correta. Diferentes tecnologias, diferentes contextos, diferentes culturas exigem diferentes ponderações. O que funciona para uma ponte pode não funcionar para um algoritmo. O que é aceitável numa sociedade pode não ser noutra. A tarefa do engenheiro reflexivo, do tecnólogo consciente, do cidadão informado, é estar atento a essas diferenças, a essas nuances. É pensar criticamente sobre o que faz, sobre o que usa, sobre o que constrói.

Esse pluralismo contextual não é relativismo, i.e., não afirma que todas as escolhas tecnológicas são igualmente justificadas. Afirma, antes, que a justificação de uma escolha tecnológica é sempre uma justificação situada, dependente de um contexto específico de valores, restrições, distribuição de poder e possibilidades técnicas. A tarefa da filosofia da engenharia e da tecnologia é precisamente tornar explícitos os pressupostos normativos das escolhas tecnológicas, identificar as alternativas que foram excluídas e perguntar em nome de quem e para quem cada escolha foi feita.

A tecnologia não é determinada pelo destino. Ela não segue um curso inevitável, construído por forças que não podemos controlar. A tecnologia é uma escolha. É resultado de decisões que tomamos, como sociedade, como empresas, como indivíduos, sobre o que desenvolver, como desenvolver, para quem desenvolver. E como toda escolha, envolve responsabilidade. Responsabilidade pelos efeitos que prevemos e pelos que não prevemos. Pelos que queremos e pelos que não queremos. Pelos que beneficiam alguns e pelos prejudicam outros. Assumir essa responsabilidade é o primeiro passo para construir uma tecnologia que mereça o nome de humana.

11

Filosofia da Ciência da Computação

1. Por que uma Filosofia da Ciência da Computação?

Quando um programador escreve uma linha de código, ele está fazendo algo fascinante. Está dando instruções a uma máquina, claro. Mas está também, sem talvez se dar conta, fazendo afirmações sobre o mundo. Dizendo o que é possível, o que não é. Delimitando fronteiras entre o que pode ser computado e o que não pode. Escolhendo representar a informação de um jeito e não de outro. Essas escolhas, aparentemente técnicas, carregam questões filosóficas profundas.

Pense num algoritmo de busca. Ele decide o que você vê primeiro, o que fica escondido na página dois, o que nem aparece. Ele define a relevância, a importância, a prioridade. Quem projetou esse algoritmo fez escolhas sobre o que conta como relevante. Essas escolhas não são neutras. Refletem valores, interesses, visões de mundo. O algoritmo pode favorecer sites mais populares, ou mais recentes, ou mais caros, ou mais alinhados com certo perfil de usuário. Cada escolha tem consequências.

A não neutralidade dos algoritmos de busca pode ser articulada ao conceito de "infraestrutura de conhecimento" (*knowledge infrastructure*) desenvolvido por Christine Borgman (2015). Sistemas técnicos que organizam e tornam acessível o conhecimento não operam como instrumentos passivos. Eles constituem artefatos sociotécnicos que incorporam valores, prioridades epistêmicas e relações de poder. A aparência de neutralidade técnica funciona, ela própria, como forma de poder, pois torna invisíveis as escolhas que o sistema materializa.

Essa opacidade não decorre apenas de fatores sociais ou institucionais. Ela se enraíza também na própria indeterminação ontológica do que se entende por programa. Um programa pode ser considerado como texto, como sequência de

caracteres, como conjunto de instruções executáveis ou como entidade abstrata. Cada uma dessas determinações corresponde a um regime distinto de acesso e de interpretação. Para o programador, trata-se do código fonte. Para o compilador, do código objeto. Para o usuário, da funcionalidade. A multiplicidade desses níveis não é contingente. Ela estrutura as condições sob as quais as decisões técnicas se tornam visíveis ou permanecem encobertas, reforçando a aparência de neutralidade dos sistemas.

A multiplicidade de perspectivas sobre o que é um programa corresponde à distinção de Colburn (2000) entre programa como objeto linguístico (uma sequência sintática), programa como objeto matemático (uma função de entradas em saídas) e programa como objeto de engenharia (um artefato que realiza uma função em um ambiente). Cada perspectiva seleciona aspectos diferentes e gera questões filosóficas distintas sobre sobre a identidade, a correção e a ontologia.

Alan Turing (1936), nos anos 1930, perguntou-se o que significa computar. Definiu uma máquina abstrata, simples, com uma fita infinita e uma cabeça de leitura e escrita. Mostrou que essa máquina, apesar da simplicidade, podia computar qualquer coisa que qualquer computador real pudesse computar. Definiu assim os limites do computável. Mostrou que há problemas que nenhuma máquina pode resolver, como o problema da parada, a saber, o problema de que não há algoritmo que diga, para qualquer programa e qualquer entrada, se esse programa vai parar ou rodar para sempre. Esse resultado é filosófico, não técnico. Não é que ainda não descobrimos o algoritmo; é que ele não pode existir. Há limites para o que a computação pode fazer. E esses limites não são práticos, são lógicos. Depois vieram os computadores reais, as linguagens de programação, a internet, a inteligência artificial. Cada avanço trouxe novas questões. O que é um algoritmo? Quando dois programas contam como o mesmo algoritmo? Como sabemos que um programa está correto? O que significa confiar num sistema que não entendemos?

A ciência da computação é uma disciplina estranha. Mistura matemática, engenharia, lógica, psicologia. Uns a veem como ciência natural (estuda fenômenos que existem independentemente), outros como ciência formal (como a matemática), outros como engenharia (cria artefactos). Provavelmente é um pouco de cada. Este capítulo é uma tentativa de pensar sobre tudo isso. Vamos

começar pela metafísica, pelo que é um algoritmo, um programa, uma computação. Depois vamos para a epistemologia, pelo que podemos conhecer sobre sistemas computacionais, como sabemos que um programa está correto, como confiamos em algoritmos opacos. E vamos terminar na ética, com as questões sobre viés algorítmico, privacidade, vigilância, autonomia.

2. Metafísica da Ciência da Computação

O que é computação? A pergunta parece simples, mas não é. Computação acontece em computadores, óbvio. Mas também acontece em outros lugares? Uma calculadora mecânica computa? Um ábaco? O cérebro humano? O universo? Depende do que entendemos por computação. Alan Turing (1936), em 1936, deu uma resposta influente. Definiu uma máquina abstrata, a máquina de Turing (1936) (*Turing machine*), composta por uma fita infinita dividida em células, uma cabeça de leitura e escrita que pode mover-se pela fita, e um conjunto finito de estados e regras. A máquina lê o símbolo na célula atual, consulta as regras, escreve um novo símbolo, move a cabeça, muda de estado. Turing (1936) mostrou que essa máquina simples pode computar qualquer coisa que qualquer computador real possa computar, desde que tenha tempo e memória suficientes. É o que se chama de tese de Church-Turing (Church, 1936; Turing, 1936), i.e., a tese de que tudo o que é efetivamente computável é Turing-computável. A tese não é um teorema matemático; não pode ser provada, porque "efetivamente computável" é uma noção informal. Mas é aceita por praticamente todos os especialistas. Todas as tentativas de definir computabilidade (o cálculo lambda de Church, as funções recursivas de Gödel-Kleene, as máquinas de registro) acabaram por mostrar-se equivalentes. Isso sugere que Turing (1936) capturou algo real, algo objetivo, sobre a natureza da computação.

Mas o que é uma máquina de Turing (1936)? É uma entidade matemática abstrata, não um objeto físico. Existe no mesmo sentido em que existem os números ou os conjuntos. Um computador real é uma implementação física aproximada de uma máquina de Turing (1936). Tem memória finita, não infinita; pode falhar, pode ter bugs. A relação entre a máquina abstrata e o computador concreto é a mesma que entre um círculo ideal e um círculo desenhado à mão, a saber, o segundo aproxima o primeiro, mas não é idêntico a ele.

A relação entre máquina abstrata e computador concreto é análoga, na filosofia da matemática, à relação entre estrutura e modelo. A máquina de Turing (1936) é uma estrutura abstrata que os computadores reais instanciam apenas aproximadamente. Isso recoloca o debate entre platonismo e fisicalismo no contexto da computação, i.e., a máquina de Turing (1936) existe como abstrato, então a computabilidade é uma propriedade do mundo matemático; se só existem os computadores físicos, a computabilidade é relativa aos substratos materiais disponíveis.

Isso levanta uma questão. A computação é uma propriedade intrínseca de sistemas físicos, ou é algo que atribuímos a eles através de interpretação? O filósofo Hilary Putnam (1988) argumentou que qualquer sistema físico suficientemente complexo pode ser interpretado como implementando qualquer computação finita, dado um esquema de mapeamento apropriado entre estados físicos e estados computacionais. Chamou a isso a "trivialidade da implementação". Se é assim, então dizer que um sistema "implementa" uma computação não diz nada substancial sobre a sua natureza física.

A conclusão de Putnam (1988) é perturbadora. Se qualquer coisa implementa qualquer computação, então a computação não é uma propriedade objetiva do mundo. É uma construção nossa, uma forma de descrever, não algo que exista independentemente. A maioria dos cientistas da computação rejeita essa conclusão. Eles argumentam que a interpretação não é arbitrária; ela é guiada pela intenção do designer, pela função do sistema e pelas práticas da comunidade. Um processador foi projetado para implementar certas operações; os engenheiros que o construíram tinham essa intenção; os programadores que escrevem código confiam nessa implementação. A computação é real porque está ancorada em práticas humanas, não apenas em interpretações livres. Que caminho teórico escolher aqui? Precisamos ainda de muito mais investigação.

A resposta ao argumento de Hilary Putnam (1988), que invoca a intenção do designer e as práticas da comunidade, pode ser articulada com a teoria dos fatos institucionais de John Searle (1995). A computação constitui uma função que existe em virtude da atribuição coletiva de um status funcional a sistemas físicos, do mesmo modo que o dinheiro existe em virtude de práticas sociais que atribuem uma função de troca a determinados objetos. Isso não torna a compu-

tação arbitrária, mas esclarece por que ela não se reduz a propriedades físicas intrínsecas. Ela é real em virtude de práticas, e não em virtude da natureza.

Outra questão metafísica central seria saber o que é um algoritmo. Algoritmos são entidades abstratas, como números? Ou são concretos, como programas escritos em linguagens específicas? O mesmo algoritmo pode ser expresso em Python, Java, C++, assembler. Ainda assim, reconhecemos todos como instâncias do mesmo algoritmo. Isso sugere que algoritmos são tipos, não tokens, i.e., entidades abstratas que podem ser instanciadas em diferentes linguagens e plataformas.

A distinção tipo/token aplicada a algoritmos gera uma versão computacional do problema dos universais. Se algoritmos são tipos, existem independentemente de qualquer instanciação, são abstratos no sentido platonista. Se são apenas padrões de tokens, sua identidade é convencional. Moshe Vardi (2012) e Yuri Gurevich (2000) desenvolveram formalizações alternativas da noção de algoritmo que pressupõem diferentes respostas a essa questão ontológica, com implicações diretas para patentes de software e verificação de equivalência entre programas.

Mas o que constitui a identidade de um algoritmo? Quando duas implementações contam como o mesmo algoritmo? Se uma implementa uma ordenação por inserção em Python e outra em Java, claramente são o mesmo algoritmo. Se uma implementa ordenação por inserção e outra por quicksort, são diferentes. Mas há casos cinzentos. O mesmo algoritmo pode ser implementado com estruturas de dados diferentes, com otimizações diferentes, com paradigmas diferentes (imperativo vs. funcional). Até que ponto as diferenças importam?

Yuri Gurevich (2000) propôs que os algoritmos são sequências de estados abstratos que satisfazem certas propriedades. Moshe Vardi (2012) defende que os algoritmos são essencialmente descrições de máquinas de Turing (1936). A questão não é meramente acadêmica. Tem implicações para a propriedade intelectual, por indicar até que ponto algoritmos podem ser patenteados, e para a verificação de programas, dada a dificuldade em sabermos que duas implementações realizam o mesmo algoritmo.

Um programa em Python é compilado para bytecode, que é interpretado por uma máquina virtual, que é implementada em C, que é compilado para linguagem de máquina, que é executada por circuitos digitais, que são compostos por

transístores. Cada nível abstrai detalhes do nível inferior e fornece uma interface. Essa arquitetura em camadas é central para a engenharia de software. Ela permite a realização múltipla (*multiple realizability*) de o mesmo algoritmo pode rodar em *hardwares* diferentes, em sistemas operacionais diferentes, em linguagens diferentes. Isso é análogo ao funcionalismo em filosofia da mente. Assim como um mesmo estado mental pode ser realizado por diferentes configurações neurais, um mesmo algoritmo pode ser realizado por diferentes implementações físicas.

A analogia entre a realização múltipla na computação e o funcionalismo na filosofia da mente possui um corolário importante. A mesma razão que sustenta a irredutibilidade dos estados mentais à neuroquímica, a saber, a múltipla realizabilidade, sustenta também a irredutibilidade dos algoritmos ao *hardware*. Disso decorre uma implicação ontológica relevante. Os níveis de abstração na computação não constituem apenas heurísticas de engenharia, mas indicam a existência de propriedades funcionais genuínas, distintas das propriedades físicas dos substratos que as realizam.

Mas a abstração nunca é perfeita. Detalhes de ordem inferior "vazam" para cima. O desempenho de um algoritmo depende da hierarquia de cache do processador. Bugs podem surgir de suposições implícitas sobre a semântica das operações, que variam entre implementações. Ataques de canal lateral exploram informações que vazam através do consumo de energia, da emissão eletromagnética e do tempo de execução. A independência do substrato é um ideal, não uma realidade completa.

3. Epistemologia da Ciência da Computação

Passemos ao conhecimento. Como sabemos que um programa está correto? A pergunta parece ter uma resposta objetiva, pois ou o programa satisfaz a sua especificação, ou não. Mas a epistemologia da correção de programas é surpreendentemente complexa. O método mais comum é o teste. Executamos o programa com algumas entradas e vemos se produz as saídas esperadas. Se falha, sabemos que há erro. Se passa, ganhamos confiança, mas não temos certeza. Edsger Dijkstra (1972) disse, numa frase famosa, que "testes podem revelar a presença de bugs, mas nunca a sua ausência". Um programa pode

passar em mil testes e falhar no milésimo primeiro. O espaço de entradas possíveis é grande demais para testar exaustivamente.

Para programas com entrada de 32 bits, há mais de 4 bilhões de casos possíveis, impraticável. Para sistemas concorrentes, o espaço de estados cresce exponencialmente com o número de processos, tornando o teste exaustivo completamente impossível. O teste é necessário, mas não suficiente.

A alternativa é a verificação formal por meio de provas matemáticas de que um programa satisfaz a sua especificação. Tony Hoare (1969) desenvolveu a lógica de Hoare (1969) (Hoare logic), um sistema formal para raciocinar sobre correção. Uma tripla de Hoare (1969) {P} C {Q} significa que, se a pré-condição P é verdadeira antes da execução do comando C, então a pós-condição Q será verdadeira após a execução, assumindo que C termina. A partir daí, podemos construir provas para programas inteiros. Hoje existem sistemas como Coq, Isabelle, Lean, que permitem desenvolver provas formais verificadas mecanicamente. O programador escreve o código e a especificação, e o sistema verifica se a prova está correta. Isso dá um grau de confiança muito maior do que o teste.

Mas a verificação formal tem limitações epistemológicas importantes. Primeiro, a especificação do programa precisa estar correta. Se especificarmos errado, podemos provar que o programa faz a coisa errada. A especificação é uma formalização de requisitos informais, e há sempre uma lacuna entre o que queremos e o que formalizamos. O filósofo James Fetzer (1988) argumentou que essa lacuna torna a verificação formal inadequada; a correção depende de interpretações informais da especificação formal.

A prova também assume que o compilador, o sistema operacional, o hardware e o próprio verificador de provas estão corretos. Um bug em qualquer um desses componentes pode invalidar a conclusão. Os críticos De Millo, Lipton & Perlis (1979) argumentaram que provas formais de programas carecem da "evidência social" que confere confiança às provas matemáticas. Provas matemáticas são revisadas pela comunidade, simplificadas, generalizadas, assimiladas. Provas de programas são frequentemente longas demais para revisão humana detalhada e dependem de ferramentas automáticas cuja correção pode ser questionada. Isso não torna a verificação formal inútil. Ela aumenta a confiança, mas não fornece certeza absoluta. É uma ferramenta poderosa em um

arsenal epistemológico que inclui também testes, revisão de código, análise estática, e experiência prática.

O argumento de De Millo, Lipton e Perlis (1979) sobre "evidência social" nas provas matemáticas conecta-se com a epistemologia social de Philip Kitcher (1990). A confiança em resultados científicos deriva não apenas da validade formal dos argumentos, mas da divisão de trabalho cognitivo da comunidade, múltiplos verificadores independentes que constroem sobre os resultados, os generalizam e os integram. As provas formais de programas, ao serem excessivamente longas e geradas automaticamente, podem não atingir esse patamar de validação social mesmo quando são formalmente corretas.

Além da correção, há a questão da complexidade. Um programa pode estar correto mas ser inutilmente lento. A teoria da complexidade computacional estuda os recursos (tempo, espaço) necessários para resolver problemas. A distinção entre problemas tratáveis (em P) e intratáveis (NP-completos) tem implicações epistemológicas profundas. Considere o problema da fatoração de inteiros, de dado um número, encontrar os seus fatores primos. É fácil verificar se uma solução está correta (basta multiplicar). Mas encontrar a solução é difícil. Os melhores algoritmos conhecidos são lentos para números grandes. A segurança do sistema RSA depende dessa dificuldade. Sabemos que o problema está em NP (verificável em tempo polinomial), mas não sabemos se está em P (solucionável em tempo polinomial). A conjectura $P \neq NP$, se verdadeira, estabeleceria um limite fundamental, o de haver verdades que podemos reconhecer quando apresentadas, mas não podemos descobrir sistematicamente em tempo razoável.

Há também limites mais absolutos. O problema da parada (*halting problem*), resolvido por Turing (1936), mostra que não existe algoritmo que determine, para qualquer programa e qualquer entrada, se o programa vai parar. Isso significa que certas propriedades de programas são indecidíveis em princípio. Não podemos ter ferramentas automáticas que garantidamente detectem todos os loops infinitos, todos os deadlocks, todos os bugs.

A prova de Turing (1936) da indecidibilidade do problema da parada utiliza um argumento diagonal análogo ao de Cantor para a inumerabilidade do contínuo e ao teorema da incompletude de Gödel (1931/1967). Qualquer sistema suficientemente expressivo contém enunciados verdadeiros que não podem ser

decididos por métodos internos ao sistema. A conexão entre esses três resultados, Cantor, Gödel (1931/1967), Turing (1936), constitui um dos mais profundos resultados da lógica e da matemática do século XX, estabelecendo limites estruturais ao que pode ser conhecido por métodos formais.

Esses resultados negativos são epistemicamente importantes. Saber o que não podemos saber é também conhecimento.

Passemos agora a uma questão epistemológica contemporânea urgente, sobre a opacidade dos algoritmos. Muitos sistemas de IA modernos, especialmente redes neurais profundas, são caixas-pretas. Recebem entradas, produzem saídas, mas não conseguimos explicar como chegaram àquela decisão. Um modelo de *deep learning* pode diagnosticar câncer com precisão superior a radiologistas humanos, mas não pode explicar por que classificou uma imagem particular como maligna. Proteção de Dados (GDPR) europeu estabelece um "direito à explicação" para decisões automatizadas. Mas como garantir esse direito se o próprio sistema é opaco?

O "direito à explicação" do GDPR pressupõe que explicações de sistemas de IA são possíveis e significativas, pressupostos que a pesquisa em explicabilidade tem questionado. Tim Miller (2019) argumenta que explicações em IA devem ser avaliadas não pela fidelidade ao mecanismo real do modelo (que pode ser irrepresentável em linguagem natural), mas por sua utilidade epistêmica para o destinatário da decisão. Uma explicação é boa se permite ao usuário antecipar o comportamento do sistema e contestar decisões injustas, independentemente de corresponder à causalidade interna do modelo.

Há duas abordagens principais. Uma é desenvolver técnicas de explicação *post hoc*, i.e., métodos que, depois da decisão, tentam explicar o que o modelo fez. LIME e SHAP são exemplos. Aproximam o modelo complexo localmente com modelos simples interpretáveis. A vantagem é que podem ser aplicados a qualquer modelo, mesmo caixas-pretas. A desvantagem é que as explicações são aproximações, podem ser enganosas, podem racionalizar em vez de revelar o raciocínio real.

A outra abordagem, defendida por Cynthia Rudin (2019), é projetar modelos inerentemente interpretáveis desde o início. Em vez de usar uma rede neural profunda e depois tentar explicá-la, usar um modelo mais simples (regressão logística, árvore de decisão) que já é transparente. Rudin (2019) argumenta que

para aplicações de alto risco (medicina, justiça criminal), a interpretabilidade deve ser priorizada sobre ganhos marginais de precisão. A vantagem é a transparência; a desvantagem é que modelos interpretáveis podem ser menos precisos em certas tarefas.

Há também a questão do viés epistémico. Algoritmos de recomendação, busca, moderação de conteúdo moldam o que vemos, o que lemos, o que sabemos. Decidem o que é relevante, o que é importante, o que merece atenção. Mas os seus critérios são opacos, proprietários, orientados por interesses comerciais. Um algoritmo de recomendação do YouTube não maximiza a verdade ou a diversidade de perspectivas; maximiza o tempo de visualização. Isso tem consequências epistêmicas, pois amplifica os conteúdos sensacionalistas, cria câmaras de eco e promove radicalização. O efeito epistêmico dos algoritmos de recomendação pode ser analisado com o conceito de câmara de eco (*echo chamber*) de Eli Pariser (2011) e com o de bolha de filtro (*filter bubble*). Enquanto a câmara de eco é produzida por escolhas sociais dos usuários (seguir pessoas com visões similares), a bolha de filtro é produzida pelo próprio sistema algorítmico, que seleciona conteúdo sem o usuário perceber. A distinção é filosoficamente relevante porque implica diferentes atribuições de responsabilidade e diferentes estratégias de intervenção.

O que significa "conhecer" num ambiente mediado por algoritmos? Como podemos confiar no que vemos se não sabemos como foi selecionado? Essas são questões epistemológicas novas, que a filosofia da computação precisa enfrentar.

4. Ética da Ciência da Computação

Chegamos às questões mais urgentes, aquelas que estão nos jornais todos os dias. Algoritmos que discriminam, sistemas de vigilância em massa, automação que elimina empregos, plataformas que manipulam emoções. A ciência da computação não é neutra; as escolhas dos seus praticantes têm consequências profundas.

Comecemos pelo viés algorítmico (*algorithmic bias*). Sistemas de IA treinados em dados históricos aprendem os padrões desses dados. Se os dados refletem discriminação passada, o algoritmo aprende a discriminar. Um sistema

de recrutamento treinado com currículos dos últimos dez anos pode aprender que homens são preferíveis para certas posições, se a empresa contratou mais homens no passado. Um sistema de pontuação de risco criminal pode aprender que negros têm maior probabilidade de reincidir, se os dados de prisões refletem policiamento enviesado.

O problema não é que os algoritmos sejam racistas ou sexistas no sentido humano, pois eles não têm intenções. O problema é que reproduzem e amplificam vieses existentes, dando-lhes a aparência de objetividade matemática. A decisão do algoritmo parece neutra, técnica, baseada em dados. Mas os dados não são neutros; carregam a história das desigualdades.

Cathy O'Neil (2016), em *Weapons of Math Destruction*, documenta vários casos. Sistemas de avaliação de professores que punem escolas pobres. Algoritmos de pontuação de crédito que penalizam minorias. Modelos de policiamento preditivo que criam ciclos viciosos em que mais policiamento em certas áreas gera mais prisões, que justificam mais polícia. O algoritmo confirma a sua própria profecia. O que fazer? Há várias propostas. Uma é auditar os algoritmos antes da implantação, testando o seu desempenho em diferentes grupos demográficos. Outra é exigir transparência de que as empresas divulguem os critérios usados. Outra ainda é envolver os grupos afetados no design dos sistemas. E uma última é simplesmente proibir certos usos, como reconhecimento facial em espaços públicos, ou algoritmos de pontuação de risco em decisões judiciais.

Mas há um problema matemático subjacente. A cientista da computação Alexandra Chouldechova (2017) mostrou que, exceto em casos especiais, não é possível satisfazer simultaneamente critérios de justiça comuns, a saber, paridade demográfica (taxas de seleção iguais entre grupos), igualdade de oportunidades (taxas de verdadeiro positivo iguais) e paridade preditiva (valores preditivos positivos iguais). Qualquer sistema automatizado deve fazer concessões entre diferentes concepções de justiça. Não há solução técnica pura; há escolhas normativas sobre que tipo de injustiça estamos dispostos a tolerar.

A impossibilidade de satisfazer simultaneamente os três critérios de justiça de Chouldechova (2017) é formalmente análoga ao teorema da impossibildade de Arrow na teoria da escolha social, uma vez que não existe procedimento de agregação de preferências que satisfaça simultaneamente um conjunto razoável de condições de equidade. Em ambos os casos, a impossibilidade não é contin-

gente, não é uma questão de melhor design, mas necessária, pois revela que os critérios são intrinsecamente incompatíveis, forçando escolhas normativas explícitas.

Passemos à privacidade e à vigilância. Sistemas computacionais coletam dados massivamente. Cada busca, cada clique, cada compra, cada localização, cada interação social é registada. Esses dados são usados para personalizar serviços, mas também para monitorar comportamentos, prever preferências, manipular decisões.

A filósofa Helen Nissenbaum (2010) desenvolveu a teoria da integridade contextual. Para ela, privacidade não significa controle total sobre as informações pessoais, nem sigilo absoluto. Significa que a informação flua de acordo com normas apropriadas a cada contexto. Informações médicas podem ser partilhadas com médicos, mas não com empregadores. Informações financeiras com bancos, mas não com redes sociais. Violações de privacidade ocorrem quando a informação flui para contextos onde não deveria.

Conforme dissemos, Shoshana Zuboff (2019) cunhou o "termo capitalismo de vigilância" para descrever o modelo de negócio dominante online. Plataformas como Google e Facebook extraem dados comportamentais massivos, processam-nos para prever comportamento, e vendem essas predições a anunciantes. O objetivo não é servir o usuário, mas antecipar e moldar o seu comportamento. Zuboff (2019) argumenta que isso representa uma ameaça existencial à autonomia humana. Quando o comportamento é continuamente monitorado e manipulado, deixamos de ser agentes autônomos.

Nesse mecanismo de coleta de dados, há também a questão da segurança. Sistemas mal projetados podem ser hackeados, expondo dados sensíveis de milhões de pessoas. Engenheiros de software têm responsabilidade de implementar salvaguardas adequadas, como criptografia, controle de acesso, auditoria, resposta a incidentes. Mas a segurança nunca é absoluta; é uma corrida armamentista entre atacantes e defensores. Sempre haverá vulnerabilidades. A questão é até que ponto os projetistas fizeram o possível para minimizá-las.

Outra questão ética urgente é a automação e o futuro do trabalho. Estudos como o de Frey e Osborne (2017) estimam que quase metade dos empregos nos EUA têm alta probabilidade de automação nas próximas décadas. Não são apenas empregos manufatureiros; advogados, contadores, jornalistas, radiolo-

gistas estão na mira. O argumento econômico clássico é que a automação aumenta a produtividade, reduz os custos, liberta os humanos para atividades mais criativas. Mas ignora questões distributivas. Os benefícios concentram-se nos proprietários do capital; os custos recaem sobre trabalhadores deslocados. Sem políticas redistributivas, a desigualdade pode aumentar dramaticamente.

Além da questão econômica, há uma questão filosófica sobre o papel do trabalho na vida humana. O trabalho não é só fonte de renda; é fonte de identidade, propósito e estrutura social. Se a automação eliminar o trabalho para muitos, o que farão essas pessoas? Como encontrarão propósito? Isso tem levado a propostas de renda básica universal. A ideia é desvincular a sobrevivência do emprego, permitindo que as pessoas busquem atividades significativas sem medo de passar fome.

A questão filosófica sobre o papel do trabalho na vida humana tem tratamento sistemático em Hannah Arendt (1958), que distingue *labor* (atividade biológica de reprodução da vida), *work* (fabricação de objetos duráveis que constituem o mundo) e *action* (atividade política de iniciativa e discurso entre iguais). A automação ameaça principalmente o *labor* e parte do *work*, mas não necessariamente a *action*. A questão relevante, à luz de Arendt (1958), é se as atividades que a automação deixa aos humanos constituem formas genuínas de *work* e *action*, ou meramente uma reestruturação do *labor* sob outra forma.

Críticos argumentam que a UBI (*universal basic income*) trata pessoas como consumidores passivos, não como produtores ativos. Melhor seria garantir um emprego significativo para todos através de programas públicos. O debate não é apenas econômico; é sobre a natureza da vida boa.

Há ainda a questão da autonomia e da manipulação. Plataformas digitais usam técnicas sofisticadas para manter a atenção dos usuários, maximizar o tempo de uso e influenciar comportamentos. Notificações push, rolagem infinita, recomendação personalizada, cores que estimulam, tudo é desenhado para prender. O usuário pensa que está a escolher, mas o ambiente foi projetado para que escolha o que a plataforma quer.

O filósofo Harry Frankfurt (1971) distinguia a *liberdade de escolha* da *liberdade da vontade*. Podemos ter escolhas, mas se essas escolhas são manipuladas, a nossa vontade não é verdadeiramente nossa. A questão é até que ponto a manipulação algorítmica compromete a autonomia. Se um sistema sabe mais

sobre nós do que nós mesmos, e usa esse conhecimento para nos levar a agir de certas maneiras, ainda somos agentes autônomos?

A distinção de Frankfurt (1971) entre desejos de primeira e segunda ordem é relevante. Manipulação algorítmica atua sobre os desejos de primeira ordem (o que queremos fazer imediatamente), potencialmente sem comprometer os desejos de segunda ordem (o que queremos querer). Mas se o sistema molda sistematicamente os desejos de segunda ordem, se altera as preferências sobre preferências, então a autonomia no sentido pleno de Frankfurt (1971) é comprometida. A personalização algorítmica profunda, ao construir perfis preditivos cada vez mais precisos, pode operar exatamente nesse nível de segunda ordem.

Finalmente, a questão da responsabilidade em sistemas complexos. Voltamos ao problema das muitas mãos. Um sistema de IA é desenvolvido por equipas grandes, treinado com dados massivos, implantado em contextos imprevisíveis. Quando algo corre mal, quem é o responsável? O programador que escreveu uma linha específica? O gestor que definiu os prazos? A empresa que priorizou o lucro? O utilizador que usou o sistema de forma imprevista? O próprio sistema? Andreas Matthias (2004) chamou a isto lacuna de responsabilidade (*responsibility gap*), tal como indicamos. Sistemas de aprendizado autônomos fazem coisas que ninguém programou explicitamente. O comportamento emerge do treino, não do design deliberado. Como atribuir responsabilidade quando não há um autor claro para a ação?

Alguns propõem que os próprios sistemas sejam considerados agentes morais, com personalidade jurídica limitada. Outros defendem que a responsabilidade deve ser distribuída por todos os envolvidos, através de mecanismos institucionais. Outros ainda argumentam que a lacuna é inevitável e que devemos simplesmente aceitar que certos sistemas são arriscados demais para serem implantados.

A questão da agência moral de sistemas de IA pode ser enquadrada com a distinção de Floridi e Cowls (2019) entre agente moral interativo (que responde ao ambiente e pode causar danos), agente moral autônomo (que age independentemente de instruções diretas) e agente moral completo (que possui intencionalidade, responsabilidade e autoconsciência). Os sistemas de aprendizado profundo atual são, no máximo, agentes morais interativos, causam efeitos com

relevância moral sem possuir a estrutura normativa que constituiria a agência moral plena. Isso implica que a responsabilidade deve ser atribuída aos humanos que os projetam, implantam e supervisionam.

A questão está em aberto. O que parece claro é que as categorias tradicionais de responsabilidade individual são insuficientes para lidar com a complexidade dos sistemas contemporâneos. Precisamos de novos modelos, novas instituições, novas formas de prestação de contas.

5. Considerações Finais

A ciência da computação é uma disciplina fascinante e estranha. Lida com entidades abstratas (algoritmos, programas, dados) que se realizam em sistemas físicos. Define os limites do que pode ser computado, do que pode ser conhecido. Cria artefatos que moldam as nossas vidas de formas profundas.

Metafisicamente, vimos que a computação não é uma propriedade simples. Os algoritmos são entidades abstratas, mas implementam-se em sistemas físicos. Níveis de abstração permitem lidar com a complexidade, mas a abstração nunca é perfeita. A tese de Church-Turing define o que é computável, mas a computação quântica desafia a nossa compreensão da eficiência computacional.

A menção à computação quântica como desafio à compreensão da eficiência computacional merece precisão. A computação quântica não refuta a tese de Church-Turing (os computadores quânticos computam as mesmas funções que as máquinas de Turing, 1936), mas desafia a tese de Church-Turing estendida, que equipara a eficiência computacional razoável com o tempo polinomial clássico. Se o algoritmo de Shor (1994) fatoriza inteiros em tempo polinomial quântico, e se $P \neq NP$ classicamente, então a computação quântica traça uma nova fronteira entre o eficientemente computável e o intratável.

Epistemologicamente, vimos que saber que um programa está correto é mais difícil do que parece. Testes dão confiança, mas não certeza. A verificação formal oferece garantias mais fortes, mas depende de especificações corretas e de ferramentas confiáveis. Os resultados da incompletude e da indecidibilidade mostram que há limites para o que podemos saber sobre programas. A opaci-

dade de sistemas de IA coloca novos desafios epistemológicos sobre como confiar no que não entendemos.

Eticamente, confrontamo-nos com questões urgentes. Viés algorítmico, privacidade, vigilância, automação, manipulação, responsabilidade, todas envolvem valores em conflito, interesses divergentes, *trade-offs* difíceis. Não há soluções técnicas para problemas políticos. Não há algoritmo que resolva questões de justiça. Há escolhas, e escolhas envolvem responsabilidade.

O pluralismo que emerge dessas reflexões reconhece que diferentes problemas exigem diferentes abordagens. O que serve para verificar um controlador de voo pode não servir para auditar um algoritmo de recomendação. O que é aceitável num contexto pode não ser noutro. A tarefa do cientista da computação reflexivo, do engenheiro de software consciente, do utilizador crítico, é estar atento a essas diferenças, a essas nuances, e pensar sobre o que faz, sobre o que usa, sobre o que constrói.

A computação não é dada por destino. É uma escolha. E como toda escolha, envolve responsabilidade.

12

Filosofia da Criptografia

1. Por que uma Filosofia da Criptografia?

Quando você compra algo online, digita a senha do banco, ou envia uma mensagem no WhatsApp, provavelmente não pensa no que está acontecendo nos bastidores. Mas há ali uma dança complexa de matemática, computação e protocolos que torna possível o que parece mágica, a saber, que a sua mensagem chegue apenas a quem deve chegar, que ninguém no meio do caminho consiga lê-la, que o destinatário possa ter certeza de que foi você quem enviou. Essa aparência de magia é precisamente o que a filosofia da criptografia precisa desfazer. Por trás dos protocolos, há escolhas conceituais sobre o que é segurança, sobre o que conta como prova, sobre que tipo de confiança é possível entre agentes que nunca se encontraram, escolhas que não são técnicas, mas filosóficas.

A criptografia está em toda parte. Protege segredos de estado e conversas de adolescentes. Garante que o voto eletrônico seja contado corretamente e que o seu dinheiro não seja roubado no caminho. Permite que estranhos na internet confiem uns nos outros sem nunca terem se encontrado. Sem a criptografia, a vida digital simplesmente não existiria. Mas essa ubiquidade não deve obscurecer o fato de que a criptografia não é um conjunto neutro de ferramentas, é uma infraestrutura de poder. Quem define os protocolos, quem controla as chaves e quem tem acesso às comunicações são questões políticas de primeira ordem, frequentemente disfarçadas de questões técnicas.

Mas o que é, afinal, um segredo num mundo digital? Diferente de um segredo no mundo físico, onde posso trancar um diário num cofre, a informação digital pode ser copiada infinitamente sem deixar rastros. O que significa dizer

que uma mensagem é “secreta” se qualquer pessoa que a intercepte pode copiá-la perfeitamente? A resposta não é óbvia.

A diferença entre o segredo físico e o segredo digital é ontológica, não apenas tecnológica. O segredo físico opera pela exclusão do acesso ao objeto; o segredo digital opera pela ininteligibilidade, o objeto está fisicamente acessível, mas sua estrutura informacional é inapreensível sem a chave.

E o que significa dizer que um sistema criptográfico é “seguro”? Seguro contra quem? Contra ataques de hoje ou daqui a vinte anos? Contra governos com supercomputadores ou contra hackers isolados? Contra alguém que tenta todas as chaves possíveis ou contra alguém que explora bugs na implementação? A segurança não é uma propriedade absoluta; é relacional, contextual, temporal.

Essa relacionalidade da segurança tem implicações metafísicas diretas. A segurança não nomeia uma propriedade intrínseca de um sistema, mas uma relação entre o sistema, um modelo de ameaça e um horizonte tecnológico. Avaliar a segurança de um sistema sem especificar esses parâmetros é como avaliar a velocidade de um objeto sem especificar o referencial.

Há também questões epistemológicas fascinantes. Como sabemos que um sistema é seguro? Não podemos testar todas as formas de ataque, pois não sabemos quais são. Baseamo-nos em provas matemáticas que mostram que quebrar o sistema é pelo menos tão difícil quanto resolver certos problemas matemáticos. Mas essas provas dependem de suposições sobre a dificuldade desses problemas, suposições que acreditamos serem verdadeiras, mas não podemos provar. A segurança criptográfica repousa, no fundo, sobre a frágil base da nossa ignorância. Achamos que fatorar números grandes é difícil porque muita gente tentou e ninguém conseguiu. Mas isso não é prova.

Uma das questões mais importantes da matemática e da ciência da computação, se P = NP, permanece em aberto, e sua resolução afirmativa tornaria obsoleta a maior parte da criptografia moderna.

Mas temos também as questões éticas, dado que a criptografia pode proteger dissidentes em regimes autoritários e terroristas planejando ataques. Pode garantir privacidade para cidadãos comuns e esconder crimes para criminosos. Governos querem backdoors para aceder a comunicações suspeitas; os criptógrafos argumentam que os *backdoors* enfraquecem a segurança para todos.

Quem tem razão? Como equilibrar a privacidade individual com a segurança coletiva?

O filósofo Philip Rogaway (2015), num artigo provocativo, argumentou que os criptógrafos não podem fingir neutralidade. O seu trabalho tem consequências políticas profundas. Capacita resistência à vigilância, mas também facilita crimes. Ignorar essas consequências é irresponsável. A criptografia é política. A posição de Rogaway (2015) representa uma inflexão significativa na autoconsciência da disciplina. Ela desafia o ideal de neutralidade técnica que dominou a criptografia acadêmica desde sua formalização matemática, e propõe que a responsabilidade moral dos criptógrafos é constitutiva de seu trabalho, não acidental a ele.

Este capítulo é uma tentativa de refletir sobre esses problemas. Vamos começar pela metafísica, pelo que significa "segredo" num mundo digital, pelo que é a aleatoriedade, pelo que significa confiar num protocolo em vez de confiar numa pessoa. Depois vamos para a epistemologia, por como sabemos que um sistema é seguro, pelo papel das provas e das suposições, pelo que significa "conhecer" num contexto de incerteza computacional. E vamos terminar na ética, com as questões mais difíceis sobre privacidade, vigilância, uso dual e responsabilidade.

2. Metafísica da Criptografia

O que é um segredo? Na vida quotidiana, a resposta parece simples. É algo que só algumas pessoas sabem. Se eu conto um segredo a um amigo e peço que não conte a mais ninguém, o segredo continua sendo segredo enquanto só nós dois soubermos. Mas no mundo digital, a coisa é mais complicada. A informação digital pode ser copiada perfeitamente, infinitamente, sem deixar vestígios. Se eu enviar uma mensagem criptografada, qualquer pessoa que a intercepte pode copiá-la. O que impede essa pessoa de saber o conteúdo não é a impossibilidade de copiar, mas a impossibilidade de entender o que copiou.

Um segredo digital não é algo escondido; é algo tornado ininteligível. A mensagem está ali, pública, viajando por cabos, passando por servidores, disponível para quem a quiser copiar. Mas está embaralhada, transformada por operações matemáticas, de modo que só quem tem a chave certa consegue

desembaralhar. O segredo não está na ausência da informação, mas na presença de informação ininteligível. Essa distinção é filosoficamente relevante porque implica que a criptografia não opera sobre o canal de comunicação, impedindo o acesso físico à mensagem, mas sobre a estrutura informacional da própria mensagem, ou seja, ela produz um objeto fisicamente acessível mas semanticamente opaco.

Shannon (1949), o pai da teoria da informação, definiu em 1949 o que significa segredo perfeito (*perfect secrecy*). Um sistema tem um segredo perfeito se o texto cifrado não fornece nenhuma informação sobre o texto claro, independentemente do poder computacional do adversário. Em termos probabilísticos, a probabilidade de uma mensagem ser a original, dado o texto cifrado, é exatamente a mesma que a probabilidade dessa mensagem antes de qualquer observação. O adversário não aprende absolutamente nada com a interceptação. O texto cifrado é, para todos os efeitos, ruído puro.

Shannon (1949) provou que o segredo perfeito é possível com a cifra de Vernam (também chamada *one-time pad*). A ideia é simples. A chave deve ser tão longa quanto a mensagem, perfeitamente aleatória, usada apenas uma vez. Cada bit da mensagem é combinado com o bit correspondente da chave (geralmente com a operação XOR). Se a chave é verdadeiramente aleatória e nunca reutilizada, o texto cifrado é matematicamente indistinguível de ruído aleatório. Não importa quantos computadores o adversário tenha, não importa quanto tempo passe, ele nunca conseguirá extrair qualquer informação sobre a mensagem original. O resultado de Shannon (1949) é notável precisamente porque é incondicional, não depende de nenhuma suposição sobre a dificuldade de problemas computacionais. O *one-time pad* é seguro contra adversários com poder computacional ilimitado, o que o distingue estruturalmente de todos os sistemas de criptografia moderna, cuja segurança é condicional a suposições que podem ser falsificadas.

Dessa forma, existe um sistema de criptografia absolutamente, matematicamente seguro. Mas o *one-time pad* tem um problema prático enorme, a saber, a chave precisa ser tão longa quanto a mensagem, e precisa ser compartilhada antecipadamente por um canal seguro. Para enviar uma mensagem de um gigabyte, é preciso ter trocado um gigabyte de chave aleatória antes. E a chave não pode ser reutilizada. Na prática, é impraticável para a maioria das aplica-

ções. O dilema do *one-time pad* é uma instância de um problema filosófico mais amplo. A distinção entre segurança ideal e segurança realizável. O sistema perfeito existe, mas é impraticável; os sistemas práticos são todos comprometidos. A questão de quais compromissos são aceitáveis é tanto técnica quanto ética.

A criptografia moderna abandonou a busca pelo segredo absoluto e adotou a segurança computacional (*computational security*). Um sistema é computacionalmente seguro se quebrá-lo requer recursos computacionais que excedem as capacidades práticas de qualquer adversário. Não garantimos que o adversário não possa, em princípio, descobrir a mensagem. Garantimos que, para descobri-la, ele precisaria de milhões de anos de computação com os melhores supercomputadores disponíveis, ou de uma quantidade de memória que não existe no universo conhecido.

Essa mudança de paradigma, do segredo absoluto ao segredo computacional, tem implicações filosóficas que raramente são articuladas; ela implica que "segurança" passou a ser um predicado graduado e condicional, não um predicado absoluto. Um sistema não é simplesmente seguro ou inseguro; é seguro contra adversários com determinados recursos, por determinado período, dado determinado estado do conhecimento matemático.

Essa mudança tem implicações metafísicas profundas. O segredo perfeito de Shannon (1949) é uma propriedade absoluta, atemporal, independente de tecnologia. Se um sistema tem um segredo perfeito hoje, o terá daqui a mil anos. O segredo computacional, por outro lado, é relacional e histórico. Depende do que sabemos sobre a dificuldade de certos problemas matemáticos, da tecnologia disponível para adversários, do progresso futuro da computação. Um sistema considerado seguro hoje pode tornar-se inseguro amanhã se alguém descobrir um algoritmo mais eficiente para o problema subjacente, ou se computadores quânticos se tornarem práticos.

Isto leva à questão sobre o que significa dizer que um problema é "difícil". A dificuldade computacional não é uma propriedade intrínseca dos problemas; é uma relação entre o problema e os recursos disponíveis para o resolver. O problema de fatorar números grandes é considerado difícil porque, até hoje, ninguém encontrou um algoritmo eficiente. Mas não há prova de que tal algoritmo não exista. A nossa confiança na segurança do RSA baseia-se na nossa igno-

rância coletiva. Achamos que é difícil porque muita gente inteligente tentou e falhou. Mas isso não é uma garantia matemática. A dificuldade computacional é, portanto, uma propriedade epistêmica antes de ser ontológica. Não sabemos de nenhum algoritmo eficiente para certos problemas, e interpretamos essa ignorância como evidência de que nenhum existe. É uma inversão do ônus da prova com consequências filosóficas profundas. A segurança de sistemas que protegem trilhões de dólares e comunicações de bilhões de pessoas repousa sobre uma ausência de conhecimento.

O teórico Goldreich (2001) argumenta que a segurança computacional é definida em relação a modelos precisos de computação e classes de complexidade, e captura algo objetivo sobre a dificuldade de certos problemas. Mas a sua objetividade é condicional, pois, se P = NP, toda a criptografia de chave pública desaba. E não sabemos se P = NP. A conjectura P ≠ NP é uma das questões em aberto mais importantes da matemática. O fato de que a segurança da internet repousa sobre uma conjectura não demonstrada é um dos fatos mais filosoficamente significativos da ciência contemporânea. Em outras palavras, ele ilustra que algumas das infraestruturas mais críticas da civilização moderna dependem não de certezas, mas de apostas epistemicamente bem fundamentadas. A segurança da internet repousa sobre uma conjectura.

Há também uma questão metafísica fascinante sobre a aleatoriedade. Os geradores de números aleatórios são cruciais para a criptografia. Chaves precisam ser aleatórias, senão um adversário pode explorar padrões. Mas o que significa dizer que uma sequência de bits é aleatória? A pergunta não é meramente técnica, ela toca em questões sobre o determinismo, sobre a natureza da informação e sobre a diferença entre ignorância e indeterminação genuína, questões com respostas diferentes dependendo das interpretações da mecânica quântica e da filosofia da física que se adota.

Uma resposta é que a aleatoriedade é ontológica. Certos eventos na natureza são fundamentalmente indeterminados. Na mecânica quântica, o resultado de uma medição é genuinamente aleatório; não há causa oculta. Se usarmos um gerador baseado em fenômenos quânticos, podemos obter aleatoriedade genuína. É o tipo mais forte, porque é irredutível.

A distinção entre aleatoriedade ontológica e epistêmica é paralela à distinção, na filosofia da ciência, entre probabilidade objetiva (propensão) e probabili-

dade subjetiva (grau de crença). Em termos de segurança em criptografia, a questão prática reside em determinar qual modalidade de aleatoriedade é adequada. A resposta da criptografia moderna é que o que importa não é a origem metafísica da aleatoriedade, mas sua indistinguibilidade computacional, o que desloca o problema da metafísica para a epistemologia computacional.

Na aleatoriedade epistêmica, algo é aleatório para nós porque não temos informação suficiente para o prever, embora seja deterministicamente determinado. O lançamento de uma moeda é deterministicamente determinado pelas condições iniciais, mas para nós é imprevisível na prática. Geradores baseados em ruído eletrônico ou movimentos do mouse funcionam nesse nível.

Na aleatoriedade algorítmica (complexidade de Kolmogorov, 1965), uma sequência é aleatória se não pode ser gerada por um programa mais curto do que a própria sequência. Uma sequência como 0101010101... não é aleatória porque pode ser gerada por "repita 01". Uma sequência sem nenhum padrão compressível é aleatória. O problema é que a complexidade de Kolmogorov (1965) é incomputável, pois não existe algoritmo que determine, para qualquer sequência, se ela é aleatória nesse sentido. A incomputabilidade da aleatoriedade algorítmica é filosoficamente importante porque revela um limite intrínseco do conhecimento. Não podemos, em geral, verificar se uma sequência é aleatória no sentido mais rigoroso do termo. Isso não significa que a noção seja vazia, mas que ela captura uma propriedade objetiva que excede nossas capacidades de verificação.

Para a criptografia, a aleatoriedade que importa é a indistinguibilidade computacional. Duas distribuições de probabilidade sobre sequências são computacionalmente indistinguíveis se nenhum algoritmo eficiente consegue dizer, com probabilidade significativamente melhor que 50%, de qual distribuição uma amostra veio. Um gerador pseudo-aleatório produz sequências que são computacionalmente indistinguíveis de verdadeiramente aleatórias, desde que o adversário tenha recursos limitados. A segurança depende dessa indistinguibilidade.

Outro conceito metafísico central é o de confiança distribuída. Tradicionalmente, confiança era uma relação entre pessoas ou entre pessoas e instituições. Você confiava no banco para guardar o seu dinheiro, no notário para autenticar documentos, no mensageiro para levar a carta. A criptografia de chave

pública, introduzida por Diffie e Hellman (1976), deu um passo importante; ela permitiu que duas partes se comunicassem secretamente sem nunca terem partilhado uma chave antes. A confiança deixou de estar numa pessoa para estar num algoritmo, num protocolo.

O passo mais radical veio com o Bitcoin e a *blockchain*. Satoshi Nakamoto (2008) propôs um sistema de dinheiro eletrônico peer-to-peer que não requer nenhuma autoridade central. Não há banco que mantenha o registo de transações; não há autoridade que emita moeda. O registo é mantido por uma rede distribuída de participantes, e a segurança é garantida por uma combinação de criptografia e incentivos econômicos. A pergunta filosófica seria, então, o que significa confiar num protocolo. A confiança não é uma relação que costumamos ter com algoritmos; confiamos em pessoas, em instituições. Confiar num protocolo é diferente. Talvez o que esteja em jogo não seja a confiança, mas a verificabilidade. Qualquer participante pode verificar independentemente a validade de todo o histórico de transações. Não é preciso confiar que o banco não está a fraudar; você mesmo pode verificar que os registos são consistentes com as regras do protocolo. Supostamente, a confiança seria substituída pela verificação.

Críticos como De Filippi e Wright (2018) argumentam que essa substituição é menos completa do que parece. Alguém escreveu o protocolo, e confiamos que esses programadores não inseriram vulnerabilidades. Alguém mantém o software, e confiamos que as atualizações não introduzirão problemas. Alguém controla a maioria do poder computacional, e confiamos que esses atores agirão honestamente. A confiança não desaparece; desloca-se para pontos diferentes do sistema. Mas o deslocamento é significativo, pois confiamos menos em pessoas e mais em matemática, em incentivos, em distribuição de poder. Esse deslocamento pode ser descrito, nos termos de Onora O'Neill, como uma transformação das condições de inteligibilidade da confiança. Em vez de confiarmos em agentes identificáveis cuja competência e integridade podemos avaliar diretamente, confiamos em propriedades matemáticas de protocolos e em estruturas de incentivos. Isso não elimina a confiança, mas exige tipos distintos de justificação, e exclui, na prática, a maioria dos usuários da capacidade de exercê-la.

3. Epistemologia da Criptografia

Como sabemos que um sistema criptográfico é seguro? A resposta intuitiva seria dizer que é porque temos uma prova matemática. Mas essa resposta esconde mais do que revela. A estrutura lógica das provas de segurança criptográfica é peculiar e merece exame filosófico cuidadoso. Trata-se, antes, de provas condicionais que estabelecem implicações entre a dificuldade de problemas computacionais e a segurança de sistemas, e cujas premissas são, na maioria dos casos, conjecturas não demonstradas.

O procedimento padrão é o seguinte. Primeiro, define-se um modelo de ameaça (*threat model*). O que o adversário pode fazer? Pode apenas ouvir passivamente? Pode modificar mensagens? Pode pedir para cifrar textos da sua escolha? Pode pedir para decifrar textos da sua escolha? Diferentes modelos capturam diferentes capacidades. Um sistema pode ser seguro num modelo e vulnerável noutro. A formalização do modelo de ameaça é, ela mesma, uma escolha epistemológica. Ao delimitar quais capacidades o adversário possui, determina quais ataques a prova cobre e, implicitamente, quais ficam fora do escopo da demonstração. O que não está no modelo simplesmente não foi considerado, e não foi provado inseguro.

Segundo, especifica-se formalmente o que significa segurança. Para cifras, geralmente é algo como "o adversário não pode aprender nenhuma informação sobre o texto claro a partir do texto cifrado". Para assinaturas digitais, algo como "o adversário não pode forjar uma assinatura para uma mensagem que não foi assinada pelo legítimo signatário". Filosoficamente, essas provas têm a estrutura de um *modus tollens* condicional, pois não provam que o sistema é seguro, mas que sua insegurança implicaria a resolubilidade eficiente de um problema presumivelmente difícil. A segurança é estabelecida por contrapositiva, não por demonstração direta.

Terceiro, demonstra-se que, se existir um adversário eficiente que quebra o sistema, então esse adversário pode ser usado para resolver um problema computacional que se acredita ser difícil. Por exemplo, se alguém quebrar o RSA, então essa pessoa consegue fatorar números grandes eficientemente. Como acreditamos que fatorar números grandes é difícil, concluímos que o RSA

é seguro. Esta é a estrutura de uma prova de segurança reducionista (*reductionist security proof*).

O que é importante notar é que estas provas não estabelecem segurança absoluta. Estabelecem uma relação de implicação, segundo a qual, se o problema subjacente é difícil, então o sistema é seguro. A segurança do sistema é condicional à dificuldade do problema.

Mas como sabemos que o problema subjacente é difícil? Esta é a questão epistemológica central. A verdade é que não temos provas de que problemas como fatoração de inteiros ou logaritmo discreto são fundamentalmente difíceis. Suspeitamos que sejam, porque gerações de matemáticos e cientistas da computação tentaram encontrar algoritmos eficientes e não conseguiram. Mas isso é evidência empírica, indutiva, não uma prova matemática. É perfeitamente possível que alguém descubra amanhã um algoritmo rápido para fatorar números grandes, tornando o RSA instantaneamente inseguro. O fundamento epistemológico da criptografia moderna é, portanto, de natureza conjectural. Ele repousa sobre a crença na dificuldade de certos problemas, crença que é amplamente partilhada e bem evidenciada, mas que tem o estatuto epistêmico de uma hipótese científica robusta, não o de um teorema demonstrado.

Neal Koblitz & Alfred Menezes (2007), num artigo intitulado “Another Look at Provable Security”, criticam a “cultura de provas” na criptografia acadêmica. Argumentam que as provas formais podem dar uma falsa sensação de segurança. Frequentemente dependem de modelos de ameaça que não capturam todas as formas como um sistema real pode ser atacado. Um adversário prático não vai tentar quebrar o algoritmo matemático; vai explorar bugs na implementação, ataques de canal lateral (medindo tempo de execução ou consumo de energia), erros de configuração, engenharia social para obter senhas. Nada disso é capturado pelas provas formais, mas tudo compromete a segurança real.

Essa crítica articula-se com a distinção filosófica entre segurança formal e segurança substancial. A primeira é uma propriedade de modelos matemáticos; a segunda é uma propriedade de sistemas implementados em ambientes adversariais reais. A distância entre os dois não é apenas técnica, ela reflete a inevitável lacuna entre qualquer modelo e o mundo que pretende capturar. Além disso, as reduções podem ser ineficientes. Pode ser que quebrar o sistema seja

equivalente a resolver o problema, mas a redução pode introduzir fatores de degradação enormes. O sistema pode ser seguro na prática mesmo que o problema seja mais fácil do que se pensava, porque a redução perde eficiência.

James Fetzer (1988) fez uma crítica análoga à verificação formal de programas, e os seus argumentos aplicam-se aqui. Provas formais operam em modelos idealizados do mundo; a implementação real sempre introduz complexidades que não estão no modelo. O compilador pode ter bugs, o hardware pode ter falhas, o sistema operacional pode não garantir os isolamentos esperados. Uma prova formal de segurança não garante que o sistema seja seguro no mundo real; garante apenas que o modelo idealizado é seguro.

O argumento de Fetzer (1988) tem uma dimensão epistemológica mais profunda, dado que ele revela que a verificação formal, como qualquer forma de justificação, é relativa a um conjunto de suposições de fundo que não são elas mesmas verificadas formalmente. A cadeia da justificação tem que terminar em algum ponto não justificado, em práticas, em hardware, em convenções, que ficam fora do escopo da prova. Isso não é uma crítica fatal à verificação formal, mas é um lembrete da sua finitude epistêmica.

Portanto, isso não significa que provas formais sejam inúteis. São ferramentas essenciais para entender as propriedades de algoritmos e protocolos, e para eliminar grandes classes de vulnerabilidades. Mas dão-nos um tipo de conhecimento que é condicional, relativo a suposições e modelos. Sabemos que o sistema é seguro se as suposições estiverem corretas e se a implementação corresponder ao modelo. É um conhecimento de segunda ordem, sobre relações entre proposições, não sobre fatos diretamente observáveis.

Um desenvolvimento epistemológico fascinante é o conceito de provas de conhecimento zero (*zero-knowledge proofs*). Introduzidas por Goldwasser, Micali e Rackoff (1989), permitem que alguém demonstre que sabe algo sem revelar o que sabe. O exemplo clássico é o da caverna de Ali Babá. Imagine uma caverna circular com duas entradas, A e B, e uma porta secreta no fundo que só se abre com uma senha. Alice alega conhecer a senha. Quer provar isso a Bob sem revelar a senha. O exemplo clássico, proposto por Goldwasser, Micali e Rackoff (1989) no artigo que fundou a área, é o da caverna de Ali Babá. Imagine uma caverna circular com duas entradas, A e B, e uma porta secreta no fundo que só pode ser aberta com uma senha. Alice quer convencer Bob de que co-

nhece a senha, mas sem jamais revelá-la. O protocolo se desenrola em rodadas repetidas. Alice entra na caverna sozinha, escolhendo aleatoriamente uma das duas entradas, sem que Bob veja qual. Depois que ela está dentro, Bob se posiciona na entrada e grita um dos lados (A ou B) pelo qual pede que ela saia. Se Alice realmente conhece a senha, ela pode abrir a porta secreta no fundo e, mesmo que tenha entrado pelo lado oposto, aparecer pelo lado que Bob solicitou. Se ela não conhece a senha, só conseguirá sair pelo lado pedido se, por acaso, tiver entrado exatamente por aquele mesmo lado, um evento que ocorre com 50% de chance em cada rodada. Repetindo o procedimento muitas vezes, a probabilidade de uma impostora acertar todas as rodadas por mera sorte cai exponencialmente. Após vinte rodadas, por exemplo, essa chance é menor que uma em um milhão. Bob pode então se convencer de que Alice de fato possui a senha, mas em momento algum obteve qualquer informação sobre qual é a senha. O que torna isso epistemologicamente intrigante é que desafia a noção tradicional de justificação. Normalmente, justificar uma crença envolve fornecer razões ou evidências que possam ser examinadas. Numa prova de conhecimento zero, a justificação é puramente interativa e protocolar. Bob não recebe nenhuma proposição que possa avaliar como evidência; apenas participa de um protocolo que, se seguido corretamente, o leva a um estado de convicção. A convicção é justificada, mas não por razões no sentido tradicional.

Na tradição epistemológica, esse tipo de justificação é desafiador para teorias que exigem que o agente justificado possua as razões de sua crença de forma articulável, como o exige o internalismo epistemológico. As provas de conhecimento zero sugerem que pode haver formas legítimas de conhecimento justificado que são externamente dependentes de processos confiáveis, sem estarem internamente ancoradas em razões conscientes acessíveis ao agente, uma instância de externalismo epistemológico aplicada a contextos matemáticos.

Formalmente, uma prova de conhecimento zero deve satisfazer três propriedades. A completude, i.e., se Alice conhece a senha e ambos seguem o protocolo, Bob será convencido. A solidez, i.e., se Alice não conhece a senha, não pode convencer Bob exceto com uma probabilidade desprezível. E o conhecimento zero, i.e., Bob não aprende nada além da veracidade da afirmação. Esta propriedade é formalizada pela noção de simulabilidade, de existir um simulador

eficiente que, sem acesso a Alice ou à senha, pode gerar uma transcrição do protocolo indistinguível de uma execução real.

Provas de conhecimento zero têm aplicações práticas importantes. A criptomoeda Zcash usa zk-SNARKs (um tipo de prova de conhecimento zero) para permitir transações completamente privadas. A rede blockchain precisa verificar que as transações são válidas (que não há gasto duplo, que o remetente tem fundos) sem saber quem enviou quanto para quem. As provas de conhecimento zero permitem exatamente isso, a saber, o sistema verifica a validade sem aprender o conteúdo. Isso tem implicações epistemológicas profundas. Numa blockchain tradicional como a do Bitcoin, todas as transações são públicas. Qualquer um pode ver quanto foi enviado de qual endereço. Em sistemas baseados em provas de conhecimento zero, a rede alcança consenso sobre a validade das transações sem conhecer os detalhes. É uma forma de conhecimento distribuído onde a verdade é estabelecida coletivamente, mas o conteúdo permanece privado.

Outra questão epistemológica importante é a confiança distribuída em sistemas descentralizados. Como vimos na secção anterior, sistemas como blockchain não eliminam a necessidade de confiança, mas transformam-na. A questão epistemológica é como sabermos em que confiar. Num sistema descentralizado, a autoridade epistêmica não está concentrada numa instituição central, mas distribuída por uma rede de participantes. Cada participante pode verificar independentemente as transações, mas isso requer recursos computacionais e conhecimento técnico que a maioria dos usuários não tem. Há uma tensão entre a promessa democrática da descentralização, qualquer um pode verificar, e a realidade da verificabilidade técnica. Na prática, apenas atores com recursos significativos exercem a função de verificação. Isso cria uma estrutura de autoridade epistêmica que é formal e igualitária, mas substantivamente concentrada. Na prática, a maioria dos usuários de Bitcoin não executa um nó completo que verifica todas as transações. Em vez disso, confiam em serviços de provedores, como exchanges e carteiras online. Isto reintroduz a confiança em entidades centralizadas, mas de forma diferente. A confiança não está no banco que mantém o registo, mas no serviço que fornece acesso à blockchain. E estes serviços podem ser hackeados, podem agir desonestamente, podem desaparecer.

Há também a confiança no protocolo em si. Como sabemos que o protocolo do Bitcoin é seguro? Porque foi analisado por especialistas, porque tem funcionado na prática por mais de uma década, porque os incentivos econômicos parecem alinhados. Isto é conhecimento empírico, não matemático. A história está cheia de sistemas que pareciam seguros até que alguém encontrou uma falha. O The DAO, um fundo de investimento descentralizado na plataforma Ethereum, tinha um bug no código que permitiu que um atacante drenasse milhões de dólares. O código tinha sido auditado, mas a falha passou despercebida.

A criptografia confronta-nos com essa realidade de forma particularmente aguda por uma razão estrutural, as consequências das falhas são assimétricas. Um sistema pode funcionar corretamente durante décadas e ser comprometido num único instante por uma falha não antecipada; e, ao contrário de falhas em outros domínios, as falhas criptográficas frequentemente são descobertas apenas quando já causaram danos irreversíveis. Essa assimetria é um argumento filosófico forte a favor do princípio da precaução em criptografia.

Isto leva de volta à questão da falibilidade do conhecimento humano. Não importa quão rigorosa seja a análise, sempre há a possibilidade de erro. A criptografia confronta-nos com esta realidade de forma particularmente aguda, porque as consequências dos erros podem ser enormes.

4. Ética da Criptografia

Chegamos às questões éticas. A criptografia está no centro das tensões fundamentais entre privacidade individual e segurança coletiva, entre liberdade e controle, entre direitos humanos e interesses de Estado. Não há posição neutra; escolhas sobre criptografia são escolhas políticas.

A tensão ética fundamental é entre privacidade e segurança coletiva. De um lado, defensores da privacidade argumentam que a criptografia forte é essencial para proteger direitos fundamentais em sociedades democráticas. Jornalistas precisam proteger fontes. Dissidentes em regimes autoritários precisam se comunicar sem serem detectados. Cidadãos comuns precisam ter certeza de que as suas transações bancárias, conversas privadas, dados de saúde não serão

expostos. Sem a criptografia, a vigilância em massa torna-se trivial, e a possibilidade de uma sociedade aberta e pluralista fica ameaçada.

A questão da privacidade em contextos digitais não se reduz à proteção de informações pessoais, pois é constitutiva da autonomia política. Uma sociedade em que todas as comunicações são potencialmente interceptáveis é uma sociedade em que o custo de dissentir é permanentemente elevado, o que produz autocensura mesmo na ausência de vigilância efetiva. O argumento de Foucault (1975) sobre o panóptico é pertinente, dado que não é necessário que alguém seja vigiado para que a possibilidade de vigilância altere seu comportamento.

De outro lado, os governos e as agências de segurança argumentam que a criptografia forte dificulta as investigações criminais e as de terrorismo. Se os criminosos podem se comunicar sem serem interceptados, se podem mover dinheiro sem deixar rastros, se podem planejar ataques em completo sigilo, a capacidade do Estado de proteger cidadãos fica severamente limitada. Este é o cerne das chamadas Cripto-Guerras (*Crypto Wars*), que se arrastam há décadas.

Nos anos 1990, o governo dos EUA tentou impor controles de exportação sobre software criptográfico, classificando-o como "munição". Empresas não podiam exportar software com criptografia forte sem licença. Ativistas como Phil Zimmermann, criador do PGP (*Pretty Good Privacy*), argumentaram que isso violava a liberdade de expressão, já que o software é uma forma de expressão. Zimmermann publicou o código fonte do PGP como livro impresso, porque os livros são protegidos pela Primeira Emenda e não podem ser controlados como munição. O código podia ser escaneado a partir do livro e compilado. Foi um golpe de gênio que expôs a absurdidade da posição do governo. O caso Zimmermann/PGP ilustra como questões técnicas e questões constitucionais podem ser inextricáveis. A decisão de que o software é ou não uma forma de expressão protegida é tanto jurídica quanto filosófica, e tem implicações para a regulação de toda a tecnologia digital.

O governo também propôs o Clipper Chip, um chip de criptografia com um backdoor que permitiria a agências governamentais decifrar comunicações autorizadas judicialmente. A ideia era que as chaves de decifração seriam divididas em duas partes, mantidas por duas agências diferentes, e só poderiam ser combinadas com ordem judicial. A proposta foi amplamente rejeitada por

especialistas, que apontaram que os *backdoors* inevitavelmente enfraquecem a segurança para todos e podem ser explorados por adversários. O Clipper Chip foi um fracasso retumbante.

O caso Apple/FBI expõe a estrutura do dilema do controle de acesso assimétrico com clareza rara. O FBI queria acesso exclusivo; a Apple argumentou que acesso exclusivo é tecnicamente impossível, porque toda vulnerabilidade é potencialmente disponível para qualquer adversário suficientemente capaz. A disputa não era sobre valores, ambas as partes concordavam que segurança e privacidade importam, mas sobre uma realidade técnica com consequências normativas. O debate ressurgiu com força após os ataques terroristas em Paris (2015) e San Bernardino (2015). O FBI exigiu que a Apple criasse uma versão do iOS que permitisse contornar a criptografia do iPhone de um dos terroristas. A Apple recusou, argumentando que criar um backdoor para um dispositivo significaria criar uma vulnerabilidade que poderia ser explorada em milhões de dispositivos. O FBI acabou por aceder ao iPhone por outros meios, pagando a uma empresa de segurança para encontrar uma vulnerabilidade, mas a questão de princípio permanece.

Whitfield Diffie e Susan Landau (2007) chamam a isto o problema do controle de acesso assimétrico. É fácil impedir que qualquer pessoa tenha acesso a algo; basta usar criptografia forte e não partilhar a chave. Mas é extremamente difícil permitir que apenas alguns atores autorizados tenham acesso, enquanto todos os outros são excluídos. Qualquer mecanismo que permita acesso autorizado cria uma porta que pode ser explorada por atacantes. Não há solução técnica conhecida para este problema; é uma limitação fundamental.

Em termos filosóficos, essa limitação é uma instância da impossibilidade de discriminação perfeita entre agentes em sistemas formais. Um mecanismo de acesso que funciona para um conjunto de agentes autorizados não pode, em geral, ser tornado inacessível para um conjunto não autorizado que possua recursos suficientes para simular as credenciais dos autorizados. A questão não é de intenção, mas de estrutura matemática.

Mesmo que houvesse acordo político sobre quando e como backdoors deveriam ser usados, a implementação técnica permaneceria problemática. Sistemas complexos têm bugs; backdoors podem ser descobertos e explorados; a

história mostra que sistemas projetados para serem seguros frequentemente falham.

Outra questão ética importante é a do anonimato. A criptografia permite não apenas comunicações secretas, mas também comunicações anônimas. Sistemas como Tor (The Onion Router) permitem navegar na internet sem revelar identidade ou localização. Criptomoedas como Monero e Zcash permitem transações financeiras não rastreáveis. Isto levanta questões profundas sobre o equilíbrio entre privacidade e responsabilização.

Defensores do anonimato apontam os benefícios. Por exemplo, as vítimas de violência doméstica podem buscar ajuda online sem medo. Denunciantes podem expor a corrupção sem retaliação. Minorias sexuais em países repressivos podem encontrar comunidades de apoio. Cidadãos comuns podem ler sobre assuntos sensíveis sem serem monitorados. Numa era de vigilância em massa, o anonimato pode ser uma ferramenta essencial de autodefesa. Nissenbaum (2010) argumenta que o anonimato não é simplesmente ausência de identidade, mas uma forma de controle sobre a própria apresentação. Permite que pessoas participem em diferentes contextos sociais sem que informações de um contexto vazem para outro. Isto é importante para autonomia e para a possibilidade de experimentação e mudança pessoal.

O argumento de Nissenbaum (2010) articula-se com sua teoria mais ampla da privacidade como integridade contextual (*contextual integrity*). A informação flui adequadamente quando o contexto de sua circulação corresponde às normas do contexto em que foi gerada. O anonimato é, nessa perspectiva, um mecanismo que garante que informações geradas num contexto não colidam indevidamente com outros contextos. Sua violação não é apenas uma perda de privacidade, mas uma ruptura da integridade contextual.

Críticos apontam os custos do anonimato, como ele facilitar crimes como a pornografia infantil, a extorsão, o terrorismo, dificultando a responsabilização por discursos de ódio e assédio online e permitindo que mercados ilegais floresçam na *dark web*. O exemplo mais famoso é o Silk Road, um mercado online que operava na rede Tor e usava Bitcoin, até que o FBI conseguiu identificar e prender o criador.

O caso Silk Road ilustra uma tensão estrutural não resolúvel por apelo a princípios únicos. O mesmo mecanismo que protege o dissidente político prote-

ge o traficante de órgãos. Essa indistinguibilidade técnica entre usos legítimos e ilegítimos é o que torna a regulação do anonimato tão difícil. Qualquer medida que comprometa o segundo necessariamente compromete o primeiro. Como equilibrar esses valores? Uma posição libertária diria que a liberdade individual deve prevalecer. Uma posição autoritária diria que a segurança coletiva exige transparência total. A maioria das pessoas está em algum ponto intermédio, querendo proteger a sua própria privacidade mas também querendo que criminosos sejam apanhados. O desafio é encontrar um equilíbrio que preserve o máximo de privacidade sem criar um ambiente onde o crime possa florescer sem consequências.

O problema é que esse equilíbrio não pode ser fixado de antemão por princípios gerais, visto que depende de avaliações empíricas sobre a efetividade das medidas de vigilância, sobre os padrões reais de criminalidade e sobre a robustez das instituições democráticas de controle. É um problema de filosofia política aplicada, não de ética abstrata. O que é importante notar é que a tecnologia não é neutra. Escolhas de design sobre o anonimato não são meramente técnicas; são escolhas sobre que tipo de sociedade queremos. Como Lawrence Lessig (1999) argumentou em *Code and Other Laws of Cyberspace*, o código é uma forma de lei, pois regula o comportamento tão efetivamente quanto os estatutos legais, mas de forma menos visível e menos democrática. Quando os desenvolvedores decidem que um sistema não coletará certos dados, ou que usará criptografia de ponta a ponta, ou que permitirá transações anônimas, estão a fazer escolhas políticas disfarçadas de escolhas técnicas.

Isto levanta a questão da responsabilidade dos desenvolvedores. Eles devem considerar as implicações sociais do seu trabalho? Ou devem simplesmente fazer a melhor tecnologia possível e deixar que a sociedade decida como usá-la? É uma versão do problema do uso dual que vimos no capítulo sobre engenharia. Um desenvolvimento recente que intensifica estas questões éticas é a emergência de sistemas descentralizados baseados em blockchain. Bitcoin e outras criptomoedas prometem um sistema financeiro sem bancos, sem governos, sem autoridades centrais. Isto é atraente para quem desconfia de instituições tradicionais, para quem vive em países com moedas instáveis, para quem quer fazer transações sem intermediários. Mas também cria novos problemas éticos.

A irrecuperabilidade das chaves privadas é uma instância de um problema filosófico mais amplo sobre a distribuição de riscos em sistemas sem autoridade central. A ausência de uma entidade que possa corrigir erros, o que é apresentado como liberdade, é também a ausência de uma entidade que possa prestar socorro. A escolha entre essas duas ausências é uma escolha política sobre a relação entre autonomia e proteção. Quem é responsável quando alguém perde as suas chaves privadas e nunca mais pode acessar os seus fundos? Em sistemas tradicionais, o banco pode ajudar a recuperar senhas perdidas. No Bitcoin, se você perder a chave privada, o dinheiro simplesmente desaparece, se torna inacessível. Isto pode ser visto como uma característica (ninguém pode confiscar os seus fundos) ou como um bug (nenhuma proteção contra o erro humano). Mas é certamente uma escolha de design com implicações éticas.

Quem regula quando criptomoedas são usadas para lavagem de dinheiro ou financiamento do terrorismo? Em sistemas tradicionais, bancos têm obrigações legais de reportar transações suspeitas. Em sistemas descentralizados, não há uma entidade claramente responsável. Isto torna a aplicação da lei muito mais difícil. Quem protege consumidores quando exchanges quebram ou são hackeadas? O caso do Mt. Gox, uma exchange de Bitcoin que faliu após perder centenas de milhões de dólares em bitcoins de clientes, é exemplar. Milhares de pessoas perderam as suas economias e não tinham a quem recorrer. Em sistemas financeiros tradicionais, há seguros, reguladores, mecanismos de proteção ao consumidor. Nas criptomoedas, esses mecanismos são muito mais fracos ou inexistentes.

A promessa da descentralização é eliminar a necessidade de confiança em autoridades; a realidade é que a confiança se desloca para novos pontos, como desenvolvedores de protocolos, mineradores com grande poder computacional, detentores de grandes quantidades de tokens. E estes novos pontos de confiança podem ser tão problemáticos quanto os antigos. De Filippi e Wright (2018) argumentam que a blockchain não elimina a necessidade de governança, apenas a transforma. Decisões sobre atualizações de protocolo, sobre resolução de disputas, sobre como lidar com bugs que permitiram roubos (como no caso do The DAO) precisam ser tomadas de alguma forma. Em sistemas descentralizados, essas decisões são tomadas através de processos complexos envolvendo desenvolvedores, mineradores, detentores de tokens, e frequentemente resul-

tam em divisões (*forks*) da comunidade. O caso do The DAO é ilustrativo, pois a comunidade Ethereum decidiu fazer um hard fork para reverter o roubo, efetivamente alterando o histórico da blockchain para devolver os fundos. Isto foi extremamente controverso, e uma minoria da comunidade rejeitou a decisão, continuando a blockchain original (*Ethereum Classic*). A decisão foi política, não técnica.

O caso do The DAO é filosoficamente instrutivo. A comunidade Ethereum decidiu fazer um *hard fork*, para reverter o roubo, alterando retroativamente o histórico da blockchain. Isso demonstrou que sistemas formalmente imutáveis são, na prática, governados por decisões coletivas que podem subverter suas próprias regras, revelando que a imutabilidade do código é uma convenção social, não uma necessidade técnica.

Há também questões éticas sobre o impacto ambiental da criptomoeda. O Bitcoin, por exemplo, consome tanta eletricidade quanto países inteiros, devido ao seu mecanismo de consenso baseado em prova de trabalho (*proof of work*). Isto é justificável? Os benefícios de um sistema financeiro descentralizado superam os custos ambientais? Alternativas como prova de participação (*proof of stake*) prometem ser mais eficientes, mas ainda estão em desenvolvimento e têm os seus próprios desafios.

O impacto ambiental do Bitcoin é um caso claro do problema das externalidades negativas em sistemas descentralizados. Os custos ambientais são distribuídos globalmente, enquanto os benefícios são capturados por participantes do sistema. Em sistemas centralizados, reguladores podem internalizar externalidades por meio de impostos ou de regulação; em sistemas descentralizados, essa internalização é estruturalmente muito mais difícil.

Philip Rogaway (2015), no seu artigo *The Moral Character of Cryptographic Work*, argumenta que criptógrafos não podem esconder-se atrás de uma pretensa neutralidade técnica. O trabalho deles tem consequências políticas profundas, dado que ele capacita a resistência à vigilância, mas também facilita os crimes; protege os dissidentes, mas também os terroristas; descentraliza o poder, mas também pode concentrá-lo em novas formas. Ignorar essas consequências é irresponsável.

Rogaway (2015) defende que os criptógrafos deveriam orientar o seu trabalho por valores morais explícitos, não apenas por critérios técnicos de elegância

ou eficiência. Deveriam perguntar-se quem é empoderado pelo sistema que criam, quem é prejudicado, como ele será usado em contextos reais. Esta posição é controversa nesse campo, que tradicionalmente se vê como ciência pura, mas torna-se cada vez mais difícil de ignorar num mundo onde a criptografia está tão entrelaçada com questões fundamentais de poder, liberdade e justiça.

A posição de Rogaway (2015) é uma aplicação ao campo da criptografia de um argumento mais geral da ética da tecnologia. Os criadores de tecnologias são agentes morais cujas responsabilidades não se esgotam na correção técnica de seus produtos, mas se estendem às consequências previsíveis de seu uso. Esse argumento é mais fácil de enunciar do que de operacionalizar, porque requer que criptógrafos desenvolvam competências em análise social e política que sua formação técnica geralmente não contempla, o que é, em si mesmo, um problema institucional e educacional.

5. Considerações Finais

A criptografia é uma disciplina fascinante porque está na interseção da matemática, da computação, da política e da ética. Lida com segredos, com confiança, com poder. As suas ferramentas protegem dissidentes e criminosos, garantem privacidade e dificultam investigações. Não há neutralidade possível.

Metafisicamente, vimos que o segredo digital não é a ausência de informação, mas a informação tornada ininteligível. O segredo perfeito é possível mas impraticável; a segurança computacional é prática mas também condicional, dependente de suposições não provadas sobre a dificuldade de problemas. A aleatoriedade revelou-se um conceito multifacetado, com dimensões ontológicas, epistêmicas e algorítmicas. A tentativa de substituir a confiança em autoridades pela confiança em protocolos mostrou-se menos simples do que parece, com a confiança se deslocando para novos pontos do sistema.

Epistemologicamente, vimos que as provas de segurança criptográfica são menos conclusivas do que parecem. São provas condicionais, dependentes de suposições não provadas. Operam em modelos idealizados que podem não capturar todas as formas como sistemas reais são atacados. Provas de conhecimento zero desafiam a nossa compreensão tradicional de justificação, mostrando que é possível estar justificado sem ter acesso a razões proposicional-

mente articuláveis. A confiança em sistemas descentralizados é, em grande parte, confiança empírica, baseada em evidência indutiva, não em garantias matemáticas.

Eticamente, confrontamo-nos com dilemas sem soluções fáceis. A tensão entre privacidade e segurança coletiva, entre anonimato e responsabilização, entre descentralização e governança, não tem resposta técnica, sempre envolve escolhas políticas e morais. O problema do controle de acesso assimétrico mostra que backdoors para uso legítimo criam vulnerabilidades para todos. O debate sobre o anonimato expõe concessões entre proteção de dissidentes e prevenção de crimes. A crítica de Rogaway (2015) lembra que criptógrafos não são técnicos neutros, mas agentes morais cujas decisões moldam o mundo.

O pluralismo que emerge dessas reflexões não é uma posição de impasse, mas o reconhecimento de que as escolhas sobre criptografia são escolhas políticas genuínas, que precisam de deliberação democrática, não apenas de otimização técnica. A tarefa filosófica, tornar explícitos os pressupostos, articular as tensões, clarificar os valores em jogo, é uma condição de possibilidade dessa deliberação. A criptografia não pode ser reduzida a uma única perspectiva. Diferentes paradigmas criptográficos privilegiam diferentes aspetos da segurança, e diferentes contextos exigem diferentes ênfases. A tarefa do criptógrafo reflexivo não é aplicar fórmulas prontas, mas exercer um julgamento informado pela compreensão filosófica clara dos problemas, conceitos e teorias envolvidos.

A criptografia é política porque lida com poder. Quem tem acesso à informação tem poder; quem controla as chaves tem poder; quem define os protocolos tem poder. Num mundo cada vez mais digital, a criptografia é uma das formas mais importantes de distribuir ou concentrar esse poder. E isso é uma questão que não pode ser deixada apenas para técnicos.

13

Filosofia da Ciência Cognitiva e da Mente

1. Por que uma Filosofia da Ciência Cognitiva?

Pense no que está acontecendo agora enquanto você lê esta frase. Seus olhos movem-se pela página ou tela, reconhecendo letras, agrupando-as em palavras, extraindo significado. Há atividade neural a ocorrer em tempo real, padrões elétricos e químicos a propagar-se por bilhões de neurônios. Mas há também algo mais. Há uma experiência subjetiva, uma sensação de estar aqui, lendo, pensando sobre o que lê. Há alguém que é você. Como é que processos físicos no cérebro dão origem a essa experiência íntima de ser um eu consciente?

Essa pergunta, tão simples de formular e tão difícil de responder, está no coração da filosofia da mente e da ciência cognitiva. O filósofo David Chalmers (1995) chamou-lhe o problema difícil da consciência (*hard problem of consciousness*). Não é difícil no sentido de complicado, como um problema matemático complexo. É difícil num outro sentido, o de que não sabemos sequer que tipo de explicação seria satisfatória. Podemos explicar como o cérebro processa a informação visual, como coordena movimentos, como consolida memórias. Mas explicar por que todo esse processamento é acompanhado por experiência subjetiva, por que há algo que é como ser um cérebro a processar informação, isso parece escapar aos nossos esquemas explicativos habituais. A distinção entre o que pode ser explicado funcionalmente e o que parece resistir a qualquer explicação funcional constitui, como veremos, o eixo central em torno do qual se organizam os debates contemporâneos em filosofia da mente.

A ciência cognitiva é a tentativa multidisciplinar de entender a mente. Os psicólogos estudam a percepção, a memória e o raciocínio através de experimentos comportamentais. Os neurocientistas mapeiam o cérebro e correlacio-

nam atividade neural com tarefas cognitivas. Os linguistas investigam como a linguagem é processada e adquirida. Os antropólogos examinam a variação cultural na cognição. Os cientistas da computação constroem modelos computacionais de processos mentais. Os filósofos perguntam o que tudo isso significa, que pressupostos estão implícitos, que conclusões podemos legitimamente extrair.

Cada disciplina traz as suas ferramentas, os seus métodos, as suas perguntas. Mas a colaboração nem sempre é fácil. O que conta como explicação para um neurocientista pode parecer superficial para um filósofo. O que é um bom modelo para um cientista da computação pode parecer irrelevante para um psicólogo. A filosofia da ciência cognitiva é, em grande medida, o esforço para entender essas tensões e para ver como as peças se encaixam, ou para perceber que talvez não se encaixem de todo.

Há questões metafísicas fundamentais. O que é a mente? É algo separado do corpo, como pensava Descartes (2004)? É apenas o cérebro em funcionamento? Ou talvez seja algo como um programa de computador rodando no hardware neural? A resposta que damos a estas perguntas molda o que consideramos uma explicação adequada e que métodos são apropriados para investigar a mente. Essa tensão interdisciplinar não é um defeito a ser corrigido, mas uma característica constitutiva de um campo que se debruça sobre fenômenos que resistem a abordagens unidisciplinares. Ela reflete a profundidade genuína dos problemas em jogo.

Há questões epistemológicas de, por exemplo, omo podemos conhecer mentes além da nossa. Tenho acesso direto à minha própria experiência, mas à dos outros só através do comportamento. Como posso saber que os outros têm experiências como as minhas? E como posso estudar cientificamente algo que só é acessível em primeira pessoa?

E há ainda questões éticas, como as que nos questionam sobre a responsabilidade que temos, caso outros animais tenham mentes conscientes. Se a neurociência desenvolver maneiras de manipular os nossos estados mentais, que limites devemos impor? Se um dia criarmos inteligências artificiais conscientes, que direitos terão?

Este capítulo é uma tentativa de pensar sobre essas questões. Vamos começar pela metafísica, pelas diferentes teorias sobre o que a mente é e como

se relaciona com o cérebro. Depois passaremos à epistemologia, examinando como podemos conhecer mentes, as nossas e as dos outros. E terminaremos na ética, com as questões mais prementes sobre consciência animal, manipulação cognitiva e a possibilidade de mentes artificiais. A articulação dessas três dimensões, metafísica, epistemológica e ética, não é meramente expositiva, elas se implicam mutuamente. A resposta que se dá à questão metafísica sobre o que é a mente determina o que contará como evidência epistemológica sobre a presença de mente em outros sistemas, o que por sua vez delimita o escopo das obrigações éticas que reconhecemos.

2. Metafísica da Mente

O problema mente-corpo é talvez o mais antigo e persistente da filosofia. Como se relacionam os estados mentais (pensamentos, sensações, emoções) com os estados físicos (processos neurais, atividade eletroquímica no cérebro)? São a mesma coisa? São coisas diferentes que interagem? São coisas diferentes que não interagem? A distinção entre essas três modalidades de resposta, identidade, interacionismo dualista e paralelismo, estrutura o espaço lógico das posições disponíveis (ainda que muitas outras posições existam no expectro) e serve de mapa para os debates que se seguem.

A primeira grande resposta na tradição ocidental veio de René Descartes (2004), no século XVII. Descartes (2004) defendia o que se chama dualismo de substância. Para ele, existem duas substâncias fundamentalmente distintas no universo. A *res cogitans* (substância pensante) é a mente, que não tem extensão espacial, é indivisível, e a sua essência é o pensamento. A *res extensa* (substância extensa) é o corpo e todo o mundo material, que tem extensão espacial, é divisível, e a sua essência é ocupar espaço. Mente e corpo são, portanto, coisas de naturezas completamente diferentes. Importa notar que o dualismo cartesiano não é apenas uma posição metafísica isolada, mas decorre de e reforça a epistemologia de Descartes (2004), pois a certeza do *cogito* funda o conhecimento do eu pensante antes e com mais segurança do que qualquer conhecimento do mundo exterior. A arquitetura metafísica dualista é, portanto, inseparável do projeto epistemológico das *Meditações* (Descartes, 1641/1984).

O argumento de Descartes (2004) parte da dúvida metódica. Posso duvidar da existência do meu corpo; talvez esteja a sonhar, talvez um gênio maligno me esteja a enganar. Mas não posso duvidar de que estou a pensar, porque a própria dúvida é um pensamento. O famoso *cogito ergo sum* (penso, logo existo) estabelece que a mente é conhecida com mais certeza que o corpo. Como mente e corpo têm propriedades diferentes (a mente é conhecida diretamente, o corpo indiretamente; a mente é indivisível, o corpo divisível), Descartes (2004) conclui que são substâncias distintas.

O dualismo tem a vantagem de capturar certas intuições que muitos de nós temos. Os nossos pensamentos parecem qualitativamente diferentes de processos físicos. Temos acesso privilegiado aos nossos próprios estados mentais. E a ideia de que a mente poderia sobreviver à morte do corpo, algo central para muitas tradições religiosas, parece fazer sentido se mente e corpo são coisas separadas.

Mas o dualismo enfrenta um problema devastador, conhecido como o problema da interação. Se mente e corpo são substâncias de naturezas completamente diferentes, como podem interagir causalmente? Dores causam comportamentos físicos, por exemplo, quando me queimo, retiro a mão do fogo. Drogas físicas causam estados mentais, por exemplo, a anestesia faz-me perder a consciência. Mas como algo não-físico pode causar mudanças no mundo físico?

Este problema foi articulado de forma precisa por Elisabeth da Boémia em correspondência com o próprio Descartes (2004), que não conseguiu oferecer uma resposta satisfatória. Como pode uma substância sem extensão exercer força sobre uma substância extensa? A força é uma categoria mecânica que pressupõe extensão e contato. O problema da interação não é, portanto, uma dificuldade posterior ao dualismo; é uma aporia que emerge imediatamente de suas premissas (Garber, 1983, p. 35-38). Isso violaria as leis de conservação da física. E como algo físico pode causar mudanças em algo não-físico? Não temos ideia de como isso poderia funcionar.

Descartes (2004) percebeu o problema e propôs que a interação ocorre na glândula pineal, uma pequena estrutura no cérebro. Mas isso não resolve a dificuldade conceitual. Mesmo que haja um local de interação, continuamos sem perceber como substâncias de naturezas diferentes podem influenciar-se mutu-

amente. A glândula pineal é tão física quanto qualquer outra parte do corpo; o problema persiste.

Além disso, a neurociência contemporânea fornece evidências esmagadoras de correlações precisas entre estados mentais e estados cerebrais. Lesões em áreas específicas causam défices mentais específicos. A estimulação elétrica de certas regiões produz experiências mentais específicas. A neuroimagem mostra padrões de atividade cerebral correlacionados com processos mentais. Essas correlações são muito mais naturais sob uma visão que identifica mente e cérebro do que sob o dualismo, que precisa explicar por que duas substâncias distintas andam perfeitamente correlacionadas. Por estas razões, o dualismo de substância é hoje amplamente rejeitado na ciência cognitiva. Mas isso não significa que o problema mente-corpo esteja resolvido.

O fisicalismo é a tese de que tudo o que existe é físico; não há substâncias ou propriedades não-físicas. A sua versão mais simples é a teoria da identidade tipo-tipo (*type-type identity theory*), defendida por autores como U.T. Place (1956), J.J.C. Smart (1959) e Herbert Feigl nos anos 1950 e 1960. Segundo esta teoria, tipos de estados mentais são idênticos a tipos de estados cerebrais. A dor é idêntica à estimulação de fibras-C. A crença de que está a chover é idêntica a um certo padrão de ativação neural. Esta identidade é do mesmo género que "água é H_2O" ou "calor é energia cinética molecular", i.e., identidades sintéticas descobertas empiricamente, não verdades conceituais. A analogia com as identidades científicas estabelecidas, como a redução termodinâmica da temperatura à energia cinética, foi explorada sistematicamente por Smart (1959) para mostrar que a identidade mente-cérebro não é categoria diferente das reduções bem-sucedidas que a ciência já realizou. O projeto redutivo implicado pela teoria da identidade tipo-tipo é, portanto, contínuo com a prática científica normal (Smart, 1959, p. 142-156). A teoria da identidade explica naturalmente as correlações mente-cérebro, dado que se estados mentais são estados cerebrais, então é óbvio que estão perfeitamente correlacionados. Também resolve o problema da interação, pois não há duas substâncias a interagir, apenas uma. E é parcimoniosa ontologicamente, por não postular entidades misteriosas para além do físico.

Mas a teoria enfrenta uma objeção séria, articulada por Hilary Putnam e Jerry Fodor (1968) nos anos 1960 e 1970, a saber, o argumento da *realização*

múltipla (*multiple realizability*). A ideia é que o mesmo tipo de estado mental pode ser realizado por diferentes tipos de estados físicos. A dor pode ocorrer em humanos (com fibras-C), em polvos (com sistemas neurais radicalmente diferentes), em alienígenas (com bioquímica completamente distinta), e talvez em robôs (com circuitos de silício). Se a dor pode ser realizada de múltiplas formas, então a dor não pode ser idêntica a nenhum tipo físico específico. A identidade tipo-tipo é falsa.

O argumento da realização múltipla não refuta o fisicalismo em geral; refuta apenas a versão tipo-tipo. Permanece aberta a versão token-token (identidade de ocorrências particulares), segundo a qual cada ocorrência particular de dor é idêntica a algum estado físico particular, sem que haja identidade entre tipos. No entanto, a versão *token-token* sacrifica a capacidade de generalização, que tornava a identidade tipo-tipo cientificamente interessante (Putnam, 1967, p. 228-235).

Este argumento foi extremamente influente e levou ao desenvolvimento do funcionalismo, a posição dominante na ciência cognitiva contemporânea. O funcionalismo, desenvolvido por Putnam (1967), Fodor (1968) e Lewis (1966), propõe que estados mentais são definidos não pela sua constituição física, mas pelo seu papel funcional, i.e., as suas relações causais com *inputs* sensoriais, outros estados mentais e *outputs* comportamentais.

O funcionalismo herda do behaviorismo a ênfase nas relações *input-output*, mas o supera ao postular estados mentais internos causalmente eficazes, em vez de limitar a explicação ao comportamento observável. Essa distinção é crucial, pois o funcionalismo é compatível com a psicologia cognitiva e com a existência de representações mentais, enquanto o behaviorismo as excluía por princípio (Fodor, 1968, p. 55-62).

O que é a dor segundo o funcionalismo? A dor é o estado tipicamente causado por dano tecidual, que causa a crença de que algo está errado, o desejo de evitar a causa, e comportamentos de proteção como retirar a mão do fogo. Qualquer sistema que realize esse papel funcional, independentemente de ser feito de neurónios, silício ou tubos de vácuo, está em estado de dor. A dor é, essencialmente, um papel causal, não uma substância específica. O funcionalismo resolve elegantemente o problema da realização múltipla. Explica por que estados mentais têm poderes causais, pois são definidos precisamente pelos

seus papéis causais. E harmoniza-se perfeitamente com a metáfora computacional que dominou a ciência cognitiva, de que se a mente é o software e o cérebro é o hardware, então os mesmos estados mentais (programas) podem rodar em diferentes hardwares.

Mas o funcionalismo enfrenta seus desafios próprios. O mais sério é o problema dos *qualia*. *Qualia* são as propriedades qualitativas da experiência consciente, como o vermelho que vemos, a dor que sentimos, o gosto do chocolate. O problema é que parece possível que dois sistemas sejam funcionalmente idênticos (tenham exatamente os mesmos papéis causais), mas tenham *qualia* diferentes.

O problema dos *qualia* pode ser formulado como um desafio à suficiência do funcionalismo, dado que os papéis causais capturam a estrutura relacional dos estados mentais, mas não capturam seu caráter intrínseco, o 'como é' de cada estado. Essa distinção entre propriedades estruturais-relacionais e propriedades intrínsecas foi explorada por Chalmers (1996) como um fundamento para o argumento de o espaço de possibilidade fenomenal ser irredutível ao funcional.

O experimento mental do espectro invertido é clássico. Imagine que as cores que você vê quando olha para um tomate (vermelho) são as mesmas que eu vejo quando olho para a grama (verde), e vice-versa. Funcionalmente, somos idênticos, pois ambos chamamos tomates de "vermelhos", ambos paramos em sinais vermelhos, ambos associamos vermelho à paixão e verde à esperança. Mas qualitativamente, as nossas experiências são invertidas. Se isto é possível, então o funcionalismo falha em capturar uma dimensão essencial da mente, a experiência consciente. A hipótese do espectro invertido é frequentemente atribuída à tradição empirista britânica (Locke, Hume, 1739/2000), que considerava a possibilidade de que ideias simples como cores não fossem diretamente comunicáveis entre sujeitos. Sua reformulação contemporânea como desafio ao funcionalismo deve-se a Ned Block (1978), que sistematizou a distinção entre funcionalismo e consciência fenomenal como dimensões independentes da mente (Block, 1978, p. 278-305).

Outro experimento mental, o dos zumbis filosóficos (*philosophical zombies*), leva a ideia ao extremo. Um zumbi filosófico é um ser física e funcionalmente idêntico a um humano consciente, mas que não tem experiência subjetiva algu-

ma. Comporta-se exatamente como um humano, grita quando se magoa, diz que está a sentir dor, evita situações perigosas. Mas por dentro não há "ninguém"; é puro processamento de informação sem consciência. Se zumbis são concebíveis (e David Chalmers argumenta que são), então a consciência não é capturada por descrições funcionais. O argumento dos zumbis é um argumento modal. Ele parte da concebibilidade para a possibilidade metafísica, e desta para conclusões sobre a não-identidade entre consciência e propriedades funcionais. A validade desse passo, de concebibilidade para possibilidade, é um dos pontos mais disputados na filosofia da mente contemporânea, como veremos adiante

John Searle (1980) apresentou um dos experimentos mentais mais influentes da filosofia da mente, o do Quarto Chinês. Imagine que Searle (1980) está trancado numa sala com um grande livro de regras em português. Por uma fenda, recebe papéis com símbolos chineses. Searle (1980) não entende chinês; para ele, os símbolos são apenas rabiscos sem sentido. Mas o livro de regras diz-lhe, em português/inglês, exatamente como manipular os símbolos, a saber, "quando vir o símbolo X, escreva o símbolo Y". Searle (1980) segue as regras diligentemente. Com o tempo, as suas respostas tornam-se tão boas que um falante nativo de chinês lá fora não consegue distinguir as suas respostas das de outro falante nativo.

A questão crucial seria se Searle (1980) entende chinês. Ele está, afinal, a gerar respostas perfeitamente adequadas em chinês. Mas de dentro da sala, a sua experiência é apenas de manipular símbolos sem significado. Não sabe que os símbolos que recebeu são perguntas sobre o tempo; não sabe que os símbolos que produziu são respostas inteligentes; não tem ideia do que nada significa. A analogia com a computação é direta. Searle (1980) é o processador, os símbolos são os dados, o livro de regras é o programa. Assim como Searle (1980) não entende chinês apesar de manipular símbolos corretamente, computadores não entendem nada apesar de processar símbolos corretamente. A manipulação sintática, por mais sofisticada que seja, não é suficiente para semântica genuína ou compreensão.

O argumento pode ser formalizado em quatro passos. (1) programas são puramente formais ou sintáticos, i.e., manipulam símbolos baseados nas suas formas, não nos seus significados. (2) mentes têm conteúdo semântico, i.e.,

estados mentais são sobre coisas, têm significado. (3) sintaxe sozinha não é suficiente para semântica. (4) Portanto, programas não são suficientes para mentes.

Este argumento pressupõe uma distinção clara entre sintaxe e semântica que nem todos os teóricos aceitam. Filósofos como Daniel Dennett (1991) argumentam que a semântica emerge de relações causais suficientemente ricas entre símbolos e mundo, de modo que a fronteira entre manipulação sintática e compreensão semântica é uma questão de grau, não de tipo. Mas o argumento de Searle (1980) tem a força de um contraexemplo intuitivo; mesmo que a semântica possa emergir de processos físicos, um simples "seguir regras" formal parece ser insuficiente para gerar a semântica (Searle, 1980, p. 417-424).

O experimento gerou uma literatura enorme de respostas e réplicas. Uma resposta comum é a resposta do sistema, de que, embora Searle (1980) não entenda chinês, o sistema completo (Searle + livro de regras + símbolos) entende. Searle (1980) contra-argumenta que se supormos que o indivíduo memorize o livro de regras e faça todos os cálculos mentalmente, o sistema inteiro estaria dentro da sua cabeça. Ele ainda assim não entende chinês. A compreensão não emerge magicamente quando o sistema é internalizado.

Outra resposta é a resposta do robô, de que se o sistema estivesse incorporado num robô que interagisse causalmente com o mundo (que visse objetos chineses, os tocasse, agisse sobre eles), então poderia desenvolver uma compreensão genuína. Searle (1980) responde que ainda que coloquemos o quarto chinês dentro do robô, adicionemos mais regras que correlacionam inputs sensoriais com símbolos e *outputs* motores com ações, ainda assim, tudo o que acontece dentro do robô é manipulação sintática de símbolos. A corporificação adiciona mais sintaxe, não semântica.

A resposta do robô antecipa, de certo modo, o programa da cognição incorporada (*embodied cognition*) que se desenvolveu nas décadas seguintes, com a tese de que a cognição não é o processamento de símbolos abstratos, mas um processo que envolve o corpo e o ambiente de modo constitutivo. Para defensores da cognição incorporada, a réplica de Searle (1980) à resposta do robô falha porque não leva a sério o papel constitutivo (não meramente instrumental) do corpo na cognição (Varela; Thompson; Rosch, 1991, p. 172-180).

Há também a resposta do simulador de cérebro. Suponha que o programa simule exatamente o processamento de um cérebro chinês, neurônio por neurônio. Searle (1980) propõe outro experimento mental no qual imaginamos que cada neurônio é substituído por um homúnculo que faz o seu trabalho. No limite, teríamos milhões de homúnculos, cada um a fazer computações simples, sem que nenhum deles entenda chinês. Onde está a compreensão no sistema? Se não está em nenhuma parte e o sistema é apenas o agregado das partes, como a compreensão emerge?

O debate sobre o Quarto Chinês permanece em aberto, mas o experimento estabeleceu que qualquer teoria da mente precisa explicar como a sintaxe se relaciona com a semântica.

Jackson (1982, 1986) apresentou outro experimento mental influente, o problema de Mary. Mary é uma neurocientista brilhante especializada na fisiologia da visão de cores. Sabe tudo o que há para saber sobre os processos neurais envolvidos na visão de cores. Ela conhece a estrutura da retina, os comprimentos de onda que os cones detectam, os caminhos neurais do olho ao córtex visual, os correlatos neurais de diferentes experiências cromáticas. Mas Mary vive num ambiente completamente preto e branco. Nunca viu cores; toda a sua informação vem de livros em preto e branco, e telas de computador em preto e branco.

A questão crucial é o que acontece quando Mary é libertada e vê vermelho pela primeira vez. Aprende algo novo? Jackson (1982) argumenta que a resposta intuitiva é claramente sim. Mary aprende como é ver vermelho, o *qualia* da experiência de vermelho. Mas se o fisicalismo fosse verdadeiro, se todos os fatos fossem fatos físicos, então Mary já conheceria todos os fatos antes de sair da sala. Logo, há fatos não-físicos sobre a experiência consciente.

Este argumento, publicado originalmente em 1982, é conhecido como o *argumento do conhecimento* (*knowledge argument*). Ele trabalha por exploração da intuição modal de que se há algo que Mary não sabia antes de sair, e se ela sabia todos os fatos físicos, então o que ela aprendeu não é um fato físico. A questão filosófica central é se essa intuição modal resiste ao escrutínio (Jackson, 1982, p. 127-136). O argumento de Jackson (1982) depende da intuição de que há algo distintivo sobre a experiência consciente (as suas propriedades qualitativas) que não é capturado por uma descrição física completa. Não impor-

ta quanto Mary soubesse sobre neurofisiologia, não saberia como é ver vermelho até experimentar. Esse "como é" (*what it's like*) parece irredutível a fatos físicos.

Thomas Nagel (1974) desenvolveu um argumento relacionado no seu famoso artigo *What Is It Like to Be a Bat?* Nagel (1974) observa que morcegos se ecolocalizam, percebem o mundo através de um sonar. Podemos aprender toda a neurofisiologia da ecolocalização, mas nunca saberemos como é ser um morcego se ecolocalizando. Há um caráter subjetivo da experiência que escapa à descrição física objetiva.

O argumento de Thomas Nagel (1974) possui uma dimensão metodológica relevante. Ele mostra que a objetividade científica, entendida como um ideal de descrição independente de ponto de vista, é estruturalmente insuficiente para apreender aquilo que constitui o núcleo da consciência, a saber, a sua dimensão subjetiva. Desse modo, uma ciência da consciência que pretenda ser completa deve desenvolver métodos capazes de articular a perspectiva de primeira pessoa com a perspectiva de terceira pessoa, sem reduzir uma à outra (Nagel, 1974, p. 435-450).

Se essa tese é correta, a consciência introduz uma dificuldade específica para a ciência. As demais propriedades investigadas podem ser tratadas como objetivas, no sentido de serem acessíveis à descrição em terceira pessoa. A consciência, contudo, envolve aspectos que permanecem vinculados à perspectiva de primeira pessoa. Esses aspectos não se deixam converter integralmente em termos objetivos, o que impõe um limite interno ao modelo clássico de explicação científica.

Fisicalistas propuseram várias respostas ao problema de Mary. A *resposta da habilidade*, defendida por David Lewis (1988) e Laurence Nemirow, argumenta que Mary não adquire conhecimento de novos fatos, mas novas habilidades, como a habilidade de reconhecer o vermelho, de imaginar o vermelho, de lembrar-se de vermelho. O conhecimento proposicional é distinto do *know how*. Mary ganha o segundo, não o primeiro. Logo, não há fatos não-físicos. Transposta para o problema de Mary, a distinção sugere que o que Mary ganha ao ver vermelho é da mesma categoria que aprender a andar de bicicleta, uma aquisição de capacidade, não de informação proposicional nova (Lewis, 1988, p. 277-290).

Os críticos objetam que esta resposta não captura a intuição de que Mary aprende algo proposicional, não apenas habilidades. Ela aprende que vermelho é assim e assado. Além disso, parece que Mary poderia adquirir todas essas habilidades através de implantes neurais que causassem experiências de vermelho artificialmente, sem que ela realmente visse tomates vermelhos. Intuitivamente, ela ainda aprenderia algo novo nessa situação.

Outra resposta é a do conhecimento velho sob novo modo. Mary aprende sobre o mesmo fato (o estado neural de ver vermelho) de uma nova maneira. Antes, conhecia esse fato sob um modo de apresentação teórico ou descritivo; depois, conhece sob um modo de apresentação experiencial. A analogia é com "Vésper é Fósforo", pois ambas as expressões se referem ao mesmo planeta (Vênus), mas sob diferentes modos de apresentação. Similarmente, Mary conhece o mesmo estado neural sob modos diferentes. Esta resposta recorre à distinção fregeana entre *sentido* (*Sinn*) e *referência* (*Bedeutung*). A mesma referência pode ser dada sob diferentes sentidos, sem que isso implique pluralidade de fatos. Aplicada ao problema de Mary, a estratégia é mostrar que a novidade epistêmica é do nível do modo de apresentação, não do nível do referente. A dificuldade é explicar por que o modo experiencial constitui tão radicalmente um diferente tipo de acesso ao mesmo fato (Frege, 1892/1980).

Esta resposta é promissora, mas enfrenta um desafio, pois precisa explicar por que o modo experiencial é tão epistemicamente impactante. Se é realmente o mesmo fato, por que conhecê-lo experiencialmente parece tão diferente de conhecê-lo teoricamente?

David Chalmers (1995) sistematizou estes debates e introduziu uma distinção que se tornou central, a saber, a diferença entre o *problema fácil* e o *problema difícil* da consciência. Os problemas fáceis da consciência são aqueles que dizem respeito a explicar funções cognitivas associadas à consciência, como a discriminação perceptual, a integração de informação, o controle deliberado de comportamento, a diferença entre sono e vigília, a atenção seletiva, a capacidade de reportar estados mentais. Estes são "fáceis" não porque sejam triviais (são empiricamente muito difíceis), mas porque sabemos em princípio que tipo de explicação seria satisfatória, a saber, mostrar como os processos físicos e funcionais implementam essas capacidades.

A distinção entre *problemas fáceis* e *problema difícil* não é uma distinção entre problemas mais e menos complexos, mas uma distinção qualitativa entre tipos de explicação. Os *problemas fáceis* são fáceis porque têm a estrutura dos problemas científicos bem-comportados, nos quais descrever o mecanismo é descrever o fenômeno. O *problema difícil* é difícil porque nenhuma descrição mecanicista parece, em princípio, suficiente para explicar por que o mecanismo é acompanhado de experiência. Por que não há apenas processamento de informação "no escuro"? Por que há algo que é como processar informação visual, sentir dor, pensar um pensamento? Chalmers (1995) argumenta que mesmo uma explicação funcional e física completa da discriminação, integração etc., deixaria o problema difícil intocado, dado que ainda poderíamos perguntar por que essas funções geram experiência.

Para mostrar que a consciência não é uma propriedade funcional, Chalmers (1996) usa o experimento mental dos zumbis filosóficos. Zumbis filosóficos são seres física e funcionalmente idênticos a humanos conscientes, mas sem experiência subjetiva alguma. Comportam-se, entretanto, exatamente como humanos, pois discriminam cores, relatam "ver vermelho", reagem a estímulos dolorosos evitando-os, até dizem "estou consciente". Mas não há nada que seja como ser eles; todo processamento é experiencialmente escuro.

O argumento de Chalmers (1996) pode ser resumido assim. (1) zumbis são concebíveis, i.e., podemos conceber consistentemente sistemas física e funcionalmente idênticos a nós sem consciência. (2) Se zumbis são concebíveis, então são metafisicamente possíveis. (3) se são possíveis, então a consciência não sobrevém logicamente a propriedades físicas e funcionais. (4) Logo, o fisicalismo e o funcionalismo são falsos.

O passo da concebibilidade para a possibilidade metafísica é o passo lógico mais controverso do argumento, tal como indicamos. Chalmers (1996) distingue concebibilidade *prima facie* de concebibilidade ideal; zumbis seriam concebíveis para um agente ideal com conhecimento perfeito. Críticos como Brian Loar (1990) argumentam que nossa incapacidade de ver como a consciência é idêntica a estados físicos reflete nossas limitações cognitivas (o modo como conceitos fenomenais funcionam), não uma verdade metafísica sobre a irredutibilidade (Loar, 1990, p. 597-616). A concebibilidade de zumbis parece mostrar que há uma lacuna conceitual entre a organização funcional e a consciência. Mesmo o

conhecimento completo da neurociência não implicaria logicamente fatos sobre *qualia*. Logo, a consciência não é redutível ao físico.

Os fisicalistas atacam várias premissas. Alguns negam que zumbis sejam genuinamente concebíveis. Pensamos que podemos concebê-los, mas estamos a confundir concebibilidade aparente com concebibilidade genuína. Antes da descoberta de que água é H_2O, parecia que podíamos conceber "XYZ", uma substância macroscopicamente indistinguível de água, mas não H_2O. Mas dado que água é H_2O essencialmente, XYZ não é realmente concebível, a não ser aparentemente. Similarmente, se a consciência é essencialmente certos processos físicos, zumbis não são genuinamente concebíveis.

Chalmers (1996) responde que água e H_2O é um caso especial. Identificações teóricas típicas, como água = H_2O ou calor = energia cinética, funcionam porque o *definiens* captura o papel funcional do *definiendum*. Mas a consciência não é definida funcionalmente; é definida pelo seu caráter fenomenal. A analogia, portanto, falha. Esta resposta recorre à distinção entre propriedades definidas extrinsecamente (por seus papéis causais ou relacionais) e propriedades definidas intrinsecamente (por seu caráter qualitativo intrínseco). A água pode ser redefinida funcionalmente (o que dissolve sais, mata a sede etc.) e depois identificada com H_2O; a consciência, segundo Chalmers (1995), resiste a essa redefinição funcional porque seu caráter qualitativo é dado diretamente, não via papel causal.

Outros aceitam que zumbis são concebíveis, mas negam que a concebibilidade implique a possibilidade metafísica. Podemos conceber impossibilidades. Para quem não é matemático, o último teorema de Fermat ser falso é concebível, entretanto ele é necessariamente verdadeiro. A concebibilidade é um guia falível para a possibilidade. Chalmers (1996) distingue entre concebibilidade para seres limitados como nós e concebibilidade ideal, para um ser omnisciente. Zumbis, argumenta, são concebíveis idealmente.

A questão sobre a relação entre concebibilidade e possibilidade é um debate central na metafísica modal. Saul Kripke (1972) mostrou que há verdades necessárias a posteriori, verdades que só são descobertas empiricamente, mas que, uma vez descobertas, valem necessariamente. Se a identidade mente-cérebro for uma verdade necessária a posteriori, então a concebibilidade de

zumbis seria apenas aparente, como a concebibilidade aparente de calor sem energia cinética (Kripke, 1972, p. 144-155).

Alguns aceitam a conclusão de Chalmers (1995), mas negam que isso seja problemático. O naturalismo não requer fisicalismo estrito; talvez a consciência seja uma propriedade natural irredutível, análoga a propriedades fundamentais da física como carga ou massa. Esta posição, por vezes chamada *dualismo de propriedades*, mantém que há apenas uma substância (a física), mas aceita propriedades mentais irredutíveis. O problema é que, se propriedades mentais são irredutíveis mas naturais, como interagem causalmente com propriedades físicas? Se não interagem, a consciência é um epifenômeno; existe, mas não faz nada. Isto é intuitivamente inaceitável para várias pessoas (não para mim, que sou epifenomenalista), dado que dores parecem causar comportamentos, desejos parecem motivar ações. Se interagem, como evitar a violação do fechamento causal do físico?

O problema do epifenomenismo é particularmente agudo para o dualismo de propriedades porque ele ameaça solapar a própria razão pela qual se postulam propriedades mentais irredutíveis, pois, se essas propriedades não têm poder causal, então são causalmente inertes e nada em nosso comportamento ou em nossos relatos sobre a experiência pode ser explicado por referência a elas. O que nos levaria, então, a afirmar[3] que existem?

O debate sobre zumbis tem implicações profundas para a ciência cognitiva. Se zumbis são possíveis, então os estudos de terceira pessoa (neuroimagem, experimentos comportamentais, modelos computacionais) nunca podem, nem em princípio, capturar a consciência fenomenal. Poderíamos ter uma neurociência completa do "processamento consciente" sem explicar por que há a experiência. Isto cria um dilema, a saber, ou aceitamos que a consciência fenomenal está fora do alcance da ciência, ou defendemos que zumbis não são possíveis e que a consciência é redutível ao funcional e físico.

[3] Se você tiver interesse na minha posição epifenomenalista, dê uma lida em "Fundamentality: Structures, powers, and a supervenience dualism" e em minha visão de mundo, publicada em *Trans/Form/Ação.*

3. Epistemologia da Mente

Passemos agora às questões epistemológicas. Como conhecemos as nossas próprias mentes? Como conhecemos as mentes dos outros? E como podemos estudar cientificamente algo que só é acessível em primeira pessoa? O problema das outras mentes é um dos mais antigos da filosofia. Como sei que outros seres têm mentes como a minha? Tenho acesso direto apenas à minha própria consciência. Observo comportamentos de outros corpos; eles gritam quando se magoam, sorriem quando estão felizes, dizem coisas como "estou com dor". Mas como sei que esses comportamentos são acompanhados por experiências subjetivas? Talvez sejam autômatos complexos, zumbis filosóficos, que se comportam como eu mas não têm nada dentro. O problema das outras mentes não é um problema teórico apenas, visto que ele tem implicações práticas diretas para a atribuição de status moral. Se não podemos verificar de modo conclusivo se outros seres têm experiências subjetivas, nossa base para atribuir-lhes obrigações morais fica epistemicamente fragilizada. Esta conexão entre epistemologia e ética da mente será retomada posteriormente.

Não se trata de um ceticismo radical sobre a existência de outras mentes. Trata-se antes de uma questão sobre a justificação epistêmica sobre que tipo de evidência justifica a crença em outras mentes. Qual é a estrutura lógica dessa inferência? A resposta clássica é o argumento por analogia. Observo que o meu próprio comportamento (gritar quando me firo, sorrir quando estou feliz) está correlacionado com os meus estados mentais (dor, alegria). Observo que outros corpos exibem comportamentos semelhantes. Por analogia, infiro que esses comportamentos estão correlacionados com estados mentais semelhantes.

John Stuart Mill (1865/1979) formalizou esse argumento em sua análise do conhecimento de outras mentes. A inferência é uma extensão do princípio indutivo de que causas similares produzem efeitos similares. A consciência alheia seria postulada como a causa mais plausível de comportamentos similares aos que, em nosso caso, são causados pela consciência (Mill, 1865/1979, p. 191-208).

Este argumento enfrenta objeções sérias. Ele é uma indução a partir de um único caso. Isso é estatisticamente frágil. Ele não é testável, dado que nunca posso verificar diretamente se outros realmente têm experiências subjetivas.

Terceiro, parece que o argumento deveria ser simétrico, pois, se outros têm mentes porque se comportam como eu, então objetos inanimados que não se comportam como eu não têm mentes. Mas nem todo comportamento semelhante implica mente (pense em autómatos ou reflexos), e talvez alguns sistemas com comportamento muito diferente tenham mentes, como seres com neurologia radicalmente diferente.

Na psicologia cognitiva e na filosofia, há um debate sobre os mecanismos efetivos de atribuição mental. A teoria-teoria (*theory-theory*) propõe que atribuímos estados mentais através de uma teoria implícita (ou explícita) sobre as relações causais entre os estímulos, os estados mentais e os comportamentos. Assim como os físicos inferem elétrons através de teorias sobre entidades inobserváveis, nós inferimos mentes através da "psicologia popular", uma teoria sobre causação mental. A teoria-teoria recorre à epistemologia das entidades teóricas, dizendo-nos que assim como elétrons e quarks são postulados por sua capacidade explicativa e preditiva, estados mentais são postulados por sua capacidade de dar conta do comportamento alheio de modo mais econômico do que alternativas puramente comportamentais. Esta analogia foi desenvolvida por Churchland (1981) em sua crítica eliminativista à psicologia popular (Churchland, 1981, p. 67-90).

A teoria da simulação (*simulation theory*) propõe, alternativamente, que atribuímos estados mentais simulando mentalmente estar na situação do outro. Uso a minha própria mente como modelo, e imagino o que pensaria ou sentiria na situação do outro, e projeto essas atitudes nele. Esta abordagem parece fenomenologicamente mais precisa, dado que frequentemente fazemos atribuições mentais através de empatia e simulação imaginativa, não através de teorização explícita.

A teoria da simulação ganhou apoio empírico com a descoberta dos neurônios-espelho em primatas não humanos (Rizzolatti *et al*., 1996) e em humanos, i.e., neurônios que disparam tanto quando o sujeito executa uma ação quanto quando observa outro executá-la. Embora o alcance exato dessa descoberta para a teoria da mente seja contestado, ela sugere que mecanismos de simulação motora podem estar na base da atribuição mental de baixo nível (Rizzolatti; Craighero, 2004, p. 169-192).

A evidência empírica sugere que ambos os mecanismos operam. A simulação é mais comum para casos semelhantes a nós e para atribuições de baixo nível, enquanto a teorização é mais comum para casos díspares e para atribuições de alto nível. A compreensão madura de outras mentes provavelmente integra ambas as estratégias.

O problema das outras mentes torna-se particularmente agudo quando consideramos animais não humanos e potenciais máquinas conscientes. Como decidir se um polvo, com o seu sistema nervoso radicalmente diferente do nosso, é consciente? Como decidir se um sistema de IA, com a sua arquitetura computacional completamente diferente, é consciente? Não podemos usar a semelhança comportamental simples, porque sistemas podem comportar-se de forma semelhante sem estados mentais semelhantes. Não podemos usar a semelhança estrutural simples, porque estruturas diferentes podem implementar as mesmas funções. E não podemos introspectar diretamente as suas experiências.

A teoria da informação integrada de Tononi (2004, 2008) propõe uma resposta parcial. Ela nos diz que um sistema é consciente na medida em que integra informações de modo irredutível ao conjunto de suas partes, e isso pode ser medido por uma quantidade chamada Phi (Φ). Esta proposta tem a vantagem de ser, em princípio, aplicável a qualquer sistema, independentemente de sua arquitetura. Mas enfrenta a crítica de que Phi pode ser alto em sistemas claramente não conscientes (como grades de computação) e baixo em sistemas que parecem conscientes (Tononi, 2004, p. 251-264).

Alguns autores propõem critérios baseados em marcadores de consciência, como a capacidade de reportar os próprios estados, a integração de informação, o acesso global ou a flexibilidade comportamental. Mas todos estes critérios são contestáveis. Podem ser satisfeitos sem consciência (pense em zumbis funcionais) ou violados com consciência (pense em pessoas com síndrome do encarceramento, que são conscientes mas não conseguem reportar). O problema das outras mentes para casos não humanos permanece sem solução consensual.

Há uma assimetria notável na epistemologia da mente. Pareço ter um acesso privilegiado à minha própria mente. Posso saber diretamente que estou com dor, que estou a pensar em filosofia, que estou a sentir ansiedade, de uma maneira que os outros não podem. Esta autoridade de primeira pessoa é tradicio-

nalmente considerada infalível ou pelo menos altamente confiável. O acesso privilegiado foi pressuposto não apenas pela tradição cartesiana, mas também por Brentano (1874), que defendia que os atos psíquicos são auto-evidentes pela *percepção interna* (*innere Wahrnehmung*), distinguindo-a da percepção exterior falível. A ideia de que temos um conhecimento incorrigível de nossos próprios estados mentais funcionou como âncora epistemológica em diversas tradições filosóficas (Brentano, 1874/1995, p. 91-100).

No entanto, a pesquisa empírica em psicologia cognitiva e social tem questionado a extensão dessa autoridade. Nisbett e Wilson (1977) mostraram que os auto-relatos sobre as causas dos nossos comportamentos são frequentemente confabulatórios. Em experimentos clássicos, as pessoas escolhiam um produto de uma fileira (por exemplo, meias) e depois explicavam por que tinham escolhido aquele. Quando todos os produtos eram idênticos, as pessoas ainda assim inventaram razões, como "este é mais macio" ou "esta cor é mais bonita". Estavam completamente inconscientes de que a escolha havia sido aleatória. Este resultado aponta para uma distinção que se tornou fundamental na psicologia cognitiva, a saber, a distinção entre processos implícitos (automáticos, inacessíveis à introspecção) e processos explícitos (deliberativos, acessíveis ao relato consciente). Nisbett e Wilson (1977) argumentam que os sujeitos têm acesso apenas aos produtos dos processos cognitivos, não aos processos em si, e que as teorias implícitas que constroem para explicar seus comportamentos são frequentemente confabulações plausíveis, não relatos acurados (Nisbett; Wilson, 1977, p. 231-259). Posso saber que decidi X, mas não tenho acesso introspectivo às computações neurais que levaram a essa decisão. Vieses cognitivos de todos os tipos afetam o autoconhecimento, como vieses de confirmação, dissonância cognitiva, racionalização motivada, e todos distorcem a nossa autopercepção. As pessoas tendem a superestimar as suas competências (o efeito *Dunning-Kruger*), a ignorar motivações indesejáveis, e a reconstruir as suas memórias de maneira autofavorável.

Eric Schwitzgebel (2008), num artigo intitulado *The Unreliability of Naive Introspection*, argumenta provocativamente que a introspecção é profundamente não confiável mesmo sobre a experiência consciente. Por exemplo, você tem experiência visual na periferia do campo visual? Não é claro, pois a introspecção diverge sobre se há experiência genuína nas bordas ou apenas a sensação

de que há. Outro exemplo seria se as emoções têm localização corporal. Algumas pessoas relatam que sim, outras que não; a introspecção não fornece uma resposta clara.

A tese de Schwitzgebel (2008) não é que a introspecção seja inútil, mas que ela é muito menos confiável do que a tradição filosófica assumiu. Isso tem consequências metodológicas sérias, pois argumentos filosóficos que dependem de intuições introspectivas sobre *qualia*, como os argumentos de Mary e dos zumbis, precisam ser calibrados com cautela, pois a fiabilidade dos dados fenomenológicos em que se baseiam é questionável (Schwitzgebel, 2008, p. 245-273). Se Schwitzgebel (2008) está correto, então a base de muito debate filosófico sobre consciência, i.e., as intuições introspectivas sobre *qualia*, sobre a presença ou ausência de experiência, é epistemicamente suspeita. Isto não elimina a filosofia da consciência, mas exige mais humildade sobre até que ponto as intuições introspectivas podem servir como guias confiáveis.

David Marr (1982), no seu influente livro *Vision*, propôs um *framework* para entender a explicação em ciência cognitiva que se tornou extremamente influente. Trata-se da distinção entre três níveis de análise. No nível computacional, pergunta-se o que o sistema faz e porquê. Que problema está a ser resolvido? Que restrições ambientais e que objetivos definem esse problema? Por exemplo, a visão implementa a inferência de estrutura tridimensional a partir de input bidimensional retiniano, com o objetivo de permitir a navegação e o reconhecimento de objetos. A noção de nível computacional em Marr (1982) é próxima do que Chomsky (1965) denominava o nível competência, i.e., uma descrição abstrata do que o sistema é capaz de fazer, independentemente de como o faz. A diferença é que Marr (1982) torna explícita a relação entre esse nível e a tarefa ecológica para a qual o sistema evoluiu, o que introduz uma dimensão funcional-adaptativa ausente no estruturalismo chomskiano.

No nível algorítmico, pergunta-se como o sistema faz o que faz. Que representações são usadas? Que algoritmos operam sobre essas representações? Por exemplo, o sistema visual usa representações de bordas, superfícies e objetos, processadas por algoritmos de deteção de bordas, agrupamento perceptual, e assim por diante. E no nível implementacional, pergunta-se em que substrato o sistema é implementado. Que hardware neural realiza os algorit-

mos? Por exemplo, a detecção de bordas é implementada no córtex visual primário V1 através de células simples e complexas.

Marr (1982) argumentou que os três níveis são relativamente independentes mas complementares. Uma explicação completa requer todos os três. Não se pode derivar o nível computacional da implementação, porque o mesmo hardware pode implementar diferentes algoritmos. Não se pode derivar o algoritmo da computação, porque o mesmo problema pode ser resolvido por diferentes algoritmos. E não se pode derivar a implementação do algoritmo, porque o mesmo algoritmo pode ser implementado em diferentes substratos. A independência relativa dos três níveis tem as implicações metodológicas de que ela justifica a autonomia explicativa da psicologia cognitiva (nível algorítmico) em relação à neurociência (nível implementacional). Isso responde à crítica reducionista radical de que a psicologia seria apenas neurociência mal compreendida. Mesmo uma neurociência completa deixaria em aberto questões sobre quais representações e algoritmos o sistema utiliza (Marr, 1982, p. 24-28).

Este *framework* tem sido extremamente influente, mas também contestado. Objetores argumentam que os níveis não são tão independentes quanto Marr (1982) sugeriu. Restrições implementacionais afetam que algoritmos são viáveis. Escolhas algorítmicas afetam que computações são possíveis. Além disso, alguns processos cognitivos podem não admitir uma decomposição clara nos três níveis. Processos emergentes, dinâmicos ou corporificados podem não ter uma descrição computacional bem definida separada da implementação.

A crítica à tricotomia de Marr (1982) converge com o programa da cognição dinâmica (*dynamical cognition*), desenvolvido por van Gelder (1995) e outros, que argumenta que sistemas cognitivos são melhor descritos como sistemas dinâmicos acoplados ao ambiente do que como sistemas de processamento de símbolos. Nessa perspectiva, a distinção entre níveis computacional e implementacional dissolve-se, pois o comportamento cognitivo emerge da dinâmica do acoplamento sistema-ambiente, não de algoritmos implementados em substratos (Van Gelder, 1995, p. 345-381). Ainda assim, o *framework* de Marr (1982) permanece um ponto de partida útil para pensar sobre a explicação multinível em ciência cognitiva.

Joseph Levine (1983) articulou um problema epistemológico que muitos consideram distinto dos problemas metafísicos sobre consciência, o *explanatory*

gap (lacuna explicativa). Mesmo supondo que o fisicalismo seja verdadeiro (que a consciência é física), há uma lacuna explicativa entre as descrições físicas e a experiência fenomenal. Levine (1983) introduziu o conceito em diálogo com a teoria da identidade, não com o dualismo. O ponto é que mesmo um fisicalista convicto deveria reconhecer que a identidade dor = estimulação de fibras-C não tem a mesma transparência explicativa que água = H_2O. A questão, portanto, não é se o fisicalismo é verdadeiro, mas se ele pode ser explanatoriamente satisfatório (Levine, 1983, p. 354-361).

Comparemos duas identidades. "Água é H_2O" parece explicativamente satisfatória, pois entendemos por que a água tem as suas propriedades macroscópicas (liquidez, transparência), dado que é H_2O. Já "dor é estimulação de fibras-C" não parece igualmente satisfatória. Mesmo que entendamos toda a neurobiologia da estimulação de fibras-C, não entendemos por que ela é acompanhada por uma experiência dolorosa qualitativa. Há uma lacuna explicativa.

A assimetria entre os dois casos reside no seguinte. A relação entre H_2O e as propriedades da água pode ser entendida a partir das propriedades dos átomos de hidrogênio e oxigênio, o que confere à identidade um caráter explanatoriamente transparente. A identidade dor = fibras-C carece dessa transparência porque não sabemos por que propriedades físicas das fibras-C devem produzir dor qualitativa, em vez de nenhuma experiência ou uma experiência completamente diferente. Levine (1983) argumenta que esta lacuna é epistemológica, não metafísica. Não mostra que a dor não seja estimulação de fibras-C. Mostra apenas que a nossa compreensão atual não conecta os dois explanatoriamente. Talvez a ciência futura venha a preencher a lacuna. Ou talvez haja limites conceptuais para a explicação da consciência.

Colin McGinn (1989) radicalizou o diagnóstico de Levine (1983), ao defender que a lacuna explicativa é permanentemente intransponível para criaturas com o nosso tipo de cognição. Seríamos cognitivamente fechados para a solução do problema da consciência da mesma forma que um rato é cognitivamente fechado para a teoria da relatividade. Isso não implicaria que o problema seja insolúvel em si, mas apenas que nossas capacidades cognitivas específicas são estruturalmente inadequadas para abordá-lo (Mcginn, 1989, p. 349-366). Mas será que realmente o são?

O *explanatory gap* coloca o desafio para a ciência cognitiva de saber que tipo de explicação da consciência seria satisfatória. Alguns propõem teorias de interface, que buscam explicar relações sistemáticas entre a estrutura fenomenal e a estrutura neural, sem pretender uma redução completa. A teoria da informação integrada de Tononi (2004), por exemplo, mapeia a integração de informação (uma propriedade neural) para a unidade da experiência (uma propriedade fenomenal). Isto não explica por que a integração gera experiência, mas estabelece correlações sistemáticas.

Outros recorrem a analogias explanatórias, procurando tornar a relação mente-cérebro mais inteligível através de analogias com outros domínios. A relação entre software e hardware, entre onda e partícula, entre propriedades emergentes em sistemas complexos. Essas analogias podem não resolver o problema, mas podem torná-lo menos misterioso.

Uma abordagem alternativa é a da fenomenologia naturalizada, proposta por autores como Petitot *et al*. (1999) e Zahavi (2004), que, em vez de partir das descrições físicas e tentar derivar as descrições fenomenológicas, partem das estruturas da experiência tal como descritas pela fenomenologia (Husserl, Merleau-Ponty) e buscam sua correlação com estruturas neurais e cognitivas. Esta abordagem não pretende resolver o *explanatory gap*, mas torna mais precisa a formulação do problema ao refinar o vocabulário fenomenológico disponível (Zahavi, 2004, p. 332-340).

Outros ainda defendem uma revisão conceitual. Os nossos conceitos de "físico" e de "experiência" podem ser inadequados. Talvez uma revisão conceitual, inspirada pela ciência futura, possa dissolver a lacuna. Ou talvez sejamos tão cognitivamente fechados para a solução do problema da consciência quanto as baratas o são para a teoria das cordas. Ainda temos um problema em aberto, esperando pela nossa reflexão filosófica cada vez mais profunda, que não para na metafísica e na epistemologia, mas adentra o mundo do dever, da ética.

4. Ética da Mente

Chegamos às questões éticas, aquelas que talvez tenham as consequências mais diretas para a nossa vida prática e para a forma como nos relacionamos com outros seres e connosco próprios. A ideia de que a consciência tem

relevância moral é uma daquelas intuições que parecem quase universais. A maioria de nós acha que causar sofrimento a um ser consciente é pior do que causar o mesmo dano físico a uma pedra ou a um programa de computador. Não é a complexidade do comportamento que importa, nem a sofisticação das capacidades cognitivas, mas sim a presença de alguém que sinta o dano, que tenha uma experiência subjetiva negativa. A tese de que é a capacidade de ter experiências subjetivas, em particular de sofrer e de bem-estar, que fundamenta a relevância moral de um ser, foi defendida por Jeremy Bentham (1789/1996) já no século XVIII. A questão não é se os animais podem raciocinar ou falar, mas se podem sofrer (Bentham, 1789/1996, p. 282-283). Essa formulação antecipa o critério da senciência que domina o debate contemporâneo sobre o *status* moral dos animais. É por isso que a forma como tratamos os animais não humanos levanta questões éticas tão urgentes.

A ciência cognitiva e a neurociência têm acumulado evidências impressionantes de que muitos animais são conscientes. Não apenas primatas, mas também cetáceos como golfinhos e baleias, aves como corvos e papagaios, e até polvos, com os seus sistemas nervosos radicalmente diferentes do nosso, exibem comportamentos que sugerem experiências subjetivas. Reconhecem-se em espelhos, o que indica alguma forma de autoconsciência. Demonstram empatia e consolam outros em sofrimento. Resolvem problemas de formas flexíveis e criativas. A Declaração de Cambridge sobre a Consciência (Low *et al*., 2012), assinada por um grupo proeminente de neurocientistas, declarou que animais não humanos, incluindo todos os mamíferos, aves e muitas outras criaturas, como polvos, possuem os substratos neurológicos que geram estados de consciência. Esta declaração representa um marco no reconhecimento científico de que a consciência não é uma propriedade exclusiva dos humanos (Low *et al*., 2012). Há sistemas de vida, de socialização e comunicação muito complexos; tudo isto aponta para a presença de algo que vale a pena levar em conta moralmente.

Singer (1975), no seu livro *Animal Liberation*, argumentou que ignorar o sofrimento animal é uma forma de preconceito análoga ao racismo e ao sexismo. Chamou-lhe *especismo*. Se a capacidade de sofrer é o que realmente importa moralmente, então não há razão para dar mais peso ao sofrimento humano simplesmente porque é humano. O sofrimento de um animal deve contar tanto

quanto o sofrimento equivalente de um humano. As práticas que submetem animais a sofrimento massivo (a agricultura industrial, a experimentação laboratorial, o entretenimento) precisam de justificativas muito fortes para serem aceitáveis.

Singer (1975) opera dentro do utilitarismo dos interesses, desenvolvido a partir de Bentham (1789/1996) e Mill (1865/1979). Para ele, o que importa moralmente é a satisfação ou frustração de interesses, e a dor representa a frustração do interesse mais básico de qualquer ser senciente. A ampliação do círculo moral para além da espécie humana é, na perspectiva de Singer (1975), uma exigência da coerência do princípio utilitarista, não uma concessão sentimental (Singer, 1975, p. 1-26).

Tom Regan (1983), noutra linha de argumentação, defendeu que os animais são "sujeitos-de-uma-vida". Têm crenças, desejos, memórias, uma identidade que persiste no tempo. Como tal, têm valor intrínseco, não apenas valor instrumental para os humanos. Têm direitos que não podem ser violados mesmo que isso traga boas consequências para os outros. A sua posição é mais forte que a de Singer (1975), mas também mais restrita, pois aplica-se a mamíferos com certa complexidade cognitiva, não necessariamente a todos os seres sencientes.

A diferença entre Singer (1975) e Regan (1983) é a diferença entre consequencialismo e deontologia no campo da ética animal. Para Singer (1975), o sofrimento de um animal pode ser pesado contra os benefícios para outros; para Regan (1983), os direitos dos animais como sujeitos-de-uma-vida funcionam como restrições que não podem ser violadas pelo cálculo consequencialista. A oposição espelha o debate mais geral entre utilitarismo e ética kantiana (Regan, 1983, p. 242-248).

Os críticos destas posições argumentam que há diferenças moralmente relevantes entre humanos e outros animais. Os humanos têm capacidades para a autonomia, para o planeamento de longo prazo, para relações morais recíprocas, para a participação em instituições sociais complexas. Isso justificaria um tratamento diferenciado. Mas esta resposta enfrenta um problema incômodo. Infantes humanos, pessoas com deficiências cognitivas severas, idosos em estado avançado de demência não têm essas capacidades, e, no entanto, continuamos a atribuir-lhes pleno *status* moral. Se a capacidade X é o critério, então

esses humanos ficariam de fora. Se não é, então por que aplicar o critério apenas aos animais?

Este argumento é conhecido na literatura como o *argumento dos casos marginais* (*argument from marginal cases*), introduzido sistematicamente por Peter Singer (1975) e desenvolvido por Jeffrey McMahan (2002). A força do argumento é de consistência. Se o critério para status moral pleno é a posse de certas capacidades cognitivas, então certos humanos ficam excluídos; se esses humanos são incluídos por outras razões, essas razões precisam ser articuladas de modo que não reavaliem arbitrariamente o status dos animais (Mcmahan, 2002, p. 148-165).

Outra questão premente diz respeito às tecnologias de intervenção cognitiva. A neurociência e a farmacologia estão a desenvolver maneiras cada vez mais precisas de modificar a nossa mente. Drogas que melhoram a atenção, que apagam memórias traumáticas, que aumentam a empatia ou a confiança. Estimulação cerebral profunda para tratar depressão. Interfaces cérebro-computador que podem vir a permitir comunicação direta entre mentes e máquinas. Tudo isto levanta questões sobre o que significa ser autêntico, sobre onde traçar o limite entre tratamento e aprimoramento, sobre quem decide o que é uma mente saudável.

A distinção entre tratamento e *aprimoramento* (*enhancement*) é contestada porque pressupõe uma concepção de funcionamento normal que pode ser cultural e historicamente contingente. Norman Daniels (2000) propõe que a linha seja traçada pela noção de *funcionamento típico da espécie* (*species-typical functioning*). O tratamento restaura o funcionamento normal, aprimoramento vai além dele. Mas este critério está sujeito à objeção de que a normalidade seria normativa, não meramente descritiva (Daniels, 2000, p. 309-322).

No caso da modulação de memórias, a promessa é aliviar o sofrimento de quem vive com memórias traumáticas intratáveis. Mas as memórias, mesmo as dolorosas, são parte do que somos. Apagar uma memória não é como apagar um ficheiro; é apagar parte da história de alguém. Quem decide se essa memória deve ser apagada? O indivíduo, sob intensa pressão emocional? O médico, com base no seu julgamento profissional? Os amigos do traumatizado? A sociedade, através de protocolos e regulamentos?

No aprimoramento cognitivo, drogas como o metilfenidato são usadas por estudantes para melhorar o desempenho em exames. Isto é uma forma de trapaça? Cria desigualdades entre quem tem acesso e quem não tem? Deveria ser permitido, regulado ou proibido? E se o aprimoramento se tornar generalizado, aqueles que não o usarem ficarão em desvantagem competitiva, criando uma pressão para aderir que pouco tem de escolha livre.

O argumento da coerção estrutural, de que a pressão de conformidade pode tornar a escolha de aprimoramento não genuinamente livre, é desenvolvido por Michael Sandel (2007), que argumenta que o aprimoramento cognitivo ameaça valores como a abertura ao dado (*giftedness*) e a humildade diante dos limites humanos, além de aprofundar desigualdades já existentes. Mas críticos contestam se esses valores realmente merecem proteção institucional em sociedades pluralistas (Sandel, 2007, p. 85-100).

Há ainda a questão da privacidade mental. As tecnologias de neuroimagem estão a tornar-se cada vez mais precisas. Já é possível, em contextos laboratoriais, identificar padrões de atividade cerebral associados a pensamentos específicos, a intenções e a mentiras. Se isto se tornar prático, quem terá acesso a essa informação? As seguradoras poderiam querer saber se você tem predisposição para certas doenças mentais. Os empregadores poderiam querer saber se está a dizer a verdade numa entrevista. Os governos poderiam querer monitorar dissidentes. O conceito de *neurodireitos* foi proposto por Rafael Yuste e Sara Goering (2017) como uma extensão necessária dos direitos humanos para o domínio neural. Os autores argumentam que as neurotecnologias emergentes ameaçam dimensões da pessoa que os direitos tradicionais não foram formulados para proteger, incluindo a privacidade do estado neural, a liberdade cognitiva e a integridade da identidade psicológica ao longo do tempo (Yuste; Goering, 2017, p. 156-159). O cérebro sempre foi o último reduto da privacidade, o lugar onde os nossos pensamentos estão a salvo de olhares alheios. Se essa barreira cair, que proteções teremos? Muitos autores propõem a criação de neurodireitos (*neurorights*) específicos para lidar com estas ameaças; mas, como sempre, a implementação destes direitos é complexa e sujeita a controvérsias.

Por fim, há a questão talvez mais especulativa, mas não menos urgente sobre a possibilidade de inteligências artificiais conscientes. Se um dia construirmos sistemas de IA que tenham experiências subjetivas, que sintam prazer e

dor, que tenham interesses e preferências, que status moral terão? Serão nossos escravos digitais, disponíveis para serem usados, copiados, apagados conforme a nossa conveniência? Ou terão direitos que limitam o que podemos fazer com eles?

A questão do *status* moral de IAs conscientes conecta-se diretamente ao problema das outras mentes discutido na seção anterior. A dificuldade de atribuir consciência a sistemas radicalmente diferentes de nós é ainda maior no caso de IAs, porque não temos sequer a homologia neurobiológica que, no caso dos animais, serve como indício de compartilhamento de mecanismos de dor e prazer. O problema é agravado pelo fato de que IAs podem exibir comportamentos que simulam consciência sem que tenhamos meios de distinguir simulação de consciência genuína.

Esta não é uma questão puramente teórica. Já há debates sobre o bem-estar de animais não humanos, e esses debates baseiam-se em evidências de consciência. Se a evidência apontar para a consciência em certos sistemas de IA, a mesma lógica se aplicará. Mas a incerteza é enorme. Não sabemos como detectar consciência em sistemas radicalmente diferentes de nós, não sabemos se a consciência pode ser copiada ou multiplicada, não sabemos que tipo de arquitetura a produziria; e, sem uma discussão filosófica, não sabemos bem nem o que é a consciência.

Alguns argumentam que é melhor errar pelo lado da cautela. Assim, se há possibilidade de que um sistema seja consciente, devemos tratá-lo como se fosse, para não arriscar cometer atrocidades. Outros respondem que isso paralisaria o desenvolvimento tecnológico e que os critérios para atribuir consciência devem ser muito exigentes. O debate está longe de qualquer consenso.

O princípio da precaução aplicado à consciência artificial é defendido por autores como Eric Schwitzgebel e Mara Garza (2015), que propõem que, dada a incerteza sobre os critérios de consciência e o alto custo moral do erro (infligir sofrimento a seres conscientes), deve-se adotar uma postura cautelosa mesmo diante de evidências inconclusivas. A posição oposta, critérios exigentes antes de atribuir status moral, é advogada por quem teme que critérios frouxos paralisem o desenvolvimento tecnológico ou tornem impossível o controle de sistemas de IA avançados (Schwitzgebel; Garza, 2015, p. 98-119).

O que todas estas questões têm em comum é que nos forçam a pensar sobre o que realmente valorizamos. Não há algoritmos que decidam quem tem status moral e quem não tem. Não há equações que determinem se devemos ou não apagar uma memória traumática. Não há fórmulas que equilibrem privacidade e segurança de uma vez por todas. Há escolhas, e escolhas envolvem responsabilidade. A ciência cognitiva pode informar essas escolhas, mostrando-nos como a mente funciona, o que causa sofrimento, o que promove bem-estar. Mas a decisão final é sempre nossa.

5. Considerações Finais

A ciência cognitiva é talvez um dos projetos intelectuais mais ambiciosos da humanidade, que seria entender a nós mesmos, a nossa mente, a nossa consciência. É um projeto que envolve múltiplas disciplinas, múltiplos métodos, múltiplas perspectivas. E é um projeto que confronta questões filosóficas profundas em cada passo. A interdisciplinaridade da ciência cognitiva não é uma contingência histórica, mas uma necessidade conceitual. Os fenômenos da mente são ao mesmo tempo físicos (processos cerebrais), funcionais (computações e representações), fenomenológicos (experiências vividas), sociais (práticas e linguagem) e normativos (razões e valores). Nenhuma disciplina isolada dispõe das ferramentas para dar conta simultaneamente de todas essas dimensões.

Metafisicamente, vimos que a natureza da mente é profundamente contestada. O dualismo, o fisicalismo reducionista e o funcionalismo oferecem respostas diferentes, cada uma com as suas forças e fraquezas. Os experimentos mentais (o Quarto Chinês, o problema de Mary, os zumbis filosóficos) revelam intuições profundas sobre a consciência que qualquer teoria adequada deve acomodar. O problema difícil da consciência lembra-nos que a experiência subjetiva pode resistir a explicações puramente funcionais ou físicas.

Epistemologicamente, vimos que o conhecimento de outras mentes, humanas, animais ou artificiais, é problemático. O argumento por analogia, a teoria-teoria e a teoria da simulação oferecem diferentes relatos de como atribuímos estados mentais, mas nenhum deles elimina completamente a incerteza. A autoridade de primeira pessoa é real, mas limitada; a introspeção é menos confiável do que tradicionalmente se supôs. O *framework* de Marr (1982) fornece

uma maneira útil de pensar sobre a explicação multinível em ciência cognitiva, e o *explanatory gap* lembra-nos que a consciência apresenta desafios explicativos especiais.

Eticamente, vimos que a consciência é central para o *status* moral. Seres conscientes têm interesses que geram obrigações morais. A evidência de consciência em animais não humanos levanta questões urgentes sobre o seu tratamento. As tecnologias de manipulação cognitiva e neuroimagem ameaçam a autonomia e a privacidade mental, exigindo novas formas de proteção. E a possibilidade de consciência artificial obriga-nos a considerar, antecipadamente, que obrigações podemos ter para com os sistemas que nós mesmos criamos.

O pluralismo paradigmático na ciência cognitiva não é um relativismo teórico. Diferentes paradigmas fazem predições distintas e são sujeitos à avaliação empírica. O pluralismo significa antes que diferentes paradigmas capturam dimensões diferentes de um fenômeno suficientemente complexo para resistir a descrição unificada. A questão de quando e como integrá-los é ela mesma uma questão filosófica de primeira ordem, que não pode ser decidida apenas por dados empíricos. A tarefa do cientista cognitivo reflexivo não é aplicar fórmulas prontas, mas exercer um julgamento informado por uma compreensão filosófica clara.

A mente é o que nos torna quem somos. Compreendê-la é compreender-nos a nós mesmos. E essa é uma jornada que está apenas começando. Mas compreender a mente é também compreender os limites do compreender. Os problemas discutidos neste capítulo, o *explanatory gap*, o problema das outras mentes, a lacuna entre terceira e primeira pessoa, sugerem que a mente é o domínio em que o conhecimento humano se defronta mais diretamente com suas próprias condições de possibilidade e com seus próprios limites. Neste sentido, a filosofia da ciência cognitiva não é apenas uma reflexão sobre uma disciplina científica; é uma forma de autognose filosófica.

14

Filosofia da Robótica e da Inteligência Artificial

1. Por que uma Filosofia da Robótica e da Inteligência Artificial?

Você já conversou com um assistente virtual como a Siri ou a Alexa? Já se perguntou se aquilo que está do outro lado, respondendo às suas perguntas, é de alguma forma inteligente? Não no sentido de ter respostas programadas, mas no sentido de entender o que você está dizendo, de saber o que está fazendo. A maioria de nós acha óbvio que não, que são apenas programas, máquinas sem consciência, sem compreensão real. Mas por que achamos isso tão óbvio? E se um dia esses sistemas se tornarem muito mais sofisticados, a ponto de não conseguirmos mais distinguir suas respostas das de um humano? O que muda?

Essas questões expõem uma distinção intuitiva entre inteligência genuína e sua simulação, e isso pressupõe critérios que, examinados com rigor, revelam-se muito menos seguros do que parecem. Elaborá-los é precisamente a tarefa da filosofia da robótica e da inteligência artificial. Essas perguntas não são novas. Alan Turing (1950), um dos pais da computação, já as fazia nos anos 1950. Ele propôs o teste de se uma máquina conseguir conversar com um humano a ponto de este não conseguir distinguir se está falando com uma máquina ou com outra pessoa, então devemos dizer que a máquina pensa. O teste de Turing (1950) (*Turing test*) tornou-se uma referência, mas também alvo de muitas críticas. Afinal, imitar um humano é o mesmo que pensar? Não seria isso apenas simular inteligência, não tê-la de fato?

O teste de Turing (1950) assume que o comportamento externo é critério suficiente para a atribuição de estados mentais, uma premissa que a filosofia da mente de orientação internalista sempre contestou. Mas a alternativa de exigir que conheçamos os estados internos de um sistema para atribuir-lhe inteligên-

cia defronta-se com o problema clássico das outras mentes, que não é resolúvel nem para humanos por meios diretos.

A robótica e a inteligência artificial são campos que nos confrontam com questões filosóficas profundas porque lidam com a possibilidade de criar algo que talvez um dia possa ser considerado inteligente, consciente, ou até mesmo uma pessoa. O que é inteligência? O que é consciência? O que é compreensão? Se construirmos um robô que se comporta exatamente como um humano em todas as situações, ele será de fato consciente ou apenas uma simulação perfeita? E como poderíamos saber a diferença? Essas questões não são apenas acadêmicas; elas determinam como projetamos sistemas, como os regulamos, que direitos ou responsabilidades lhes atribuímos, e como os integramos na vida social. A filosofia da robótica não é posterior à prática técnica; é uma de suas condições de possibilidade.

Há também questões sobre a relação entre corpo e mente que a robótica torna concretas. Tradicionalmente, a filosofia da mente perguntava se a mente poderia existir independentemente do corpo. A robótica inverte a pergunta. Se um corpo artificial pode abrigar uma mente, que tipo de corpo é necessário? Um robô com rodas pode ser inteligente? Precisa de braços, pernas, sensores táteis? A cognição depende de ter um corpo que interage com o mundo de determinadas maneiras? Essas questões convergem com o programa da cognição corporificada (*embodied cognition*), que desafia a hegemonia do paradigma computacional clássico na ciência cognitiva, propondo que os processos cognitivos não estão confinados ao cérebro, mas se estendem ao corpo e ao ambiente.

E depois há as questões éticas. Carros autônomos vão ter que tomar decisões em situações de acidente; por exemplo, se um atropelamento é inevitável, quem o carro deve escolher atingir? Robôs de cuidado para idosos vão substituir o contato humano? Sistemas de IA usados em contratações, concessão de crédito, decisões judiciais podem reproduzir e amplificar preconceitos existentes. Armas autônomas podem selecionar e atacar alvos sem intervenção humana. Quem é responsável quando um sistema autônomo causa dano? O programador? O fabricante? O usuário? O próprio sistema?

A especificidade filosófica dessas questões reside no fato de que não se trata apenas de aplicar uma ética já constituída a novos casos, mas de que a

autonomia crescente dos sistemas artificiais desafia categorias morais fundamentais, como responsabilidade, intenção e agência, que foram elaboradas pressupondo agentes humanos.

Vamos começar pela metafísica, perguntando o que é um robô, o que é inteligência artificial, e como a corporificação se relaciona com a cognição. Depois passaremos à epistemologia, examinando como podemos saber se um sistema é inteligente ou consciente, e como podemos confiar em sistemas que não entendemos completamente. E terminaremos na ética, com as questões mais prementes sobre responsabilidade, viés, autonomia e o futuro do trabalho.

2. Metafísica da Robótica e da IA

O que é um robô? A palavra evoca imagens de máquinas humanoides, como os robôs da ficção científica. Mas na prática, um robô pode ser qualquer sistema físico que percebe o ambiente e age sobre ele de forma relativamente autônoma. Um aspirador de pó Roomba é um robô. Um braço mecânico numa linha de montagem é um robô. Um carro autônomo é um robô. O que todos têm em comum é a capacidade de operar no mundo físico com algum grau de autonomia, tomando decisões com base em sensores e executando ações através de atuadores.

A inteligência artificial, por outro lado, não precisa ter corpo. Um sistema de IA pode existir puramente como software, processando informação, reconhecendo padrões, tomando decisões, sem nunca interagir fisicamente com o mundo. O GPT, que gera texto, é IA. O sistema de recomendação da Netflix é IA. O algoritmo que detecta fraudes em cartões de crédito é IA. São todos programas que processam informação de maneiras sofisticadas, mas não têm corpos, não agem no mundo físico.

Essa distinção, entre sistemas situados fisicamente e sistemas puramente computacionais, é uma das mais carregadas do ponto de vista filosófico, porque ela separa duas tradições de pesquisa com pressupostos radicalmente diferentes sobre o que a inteligência é e o que ela requer. A cognição depende de ter um corpo? Ou a mente pode ser entendida como um programa de computador rodando em qualquer hardware, independentemente de ter ou não braços e pernas? Essa posição, conhecida também como funcionalismo, foi formulada

com rigor por Hilary Putnam (1975) e tornou-se durante décadas o paradigma dominante na filosofia da mente analítica. Mas enfrenta objeções sérias, tanto de dentro quanto de fora do paradigma computacional. Essa é a aposta do que se chama IA forte (*strong AI*). A ideia é que a mente é para o cérebro o que o software é para o hardware. Assim como o mesmo programa pode rodar em computadores diferentes, a mesma mente poderia, em princípio, rodar em substratos diferentes, sejam neurônios biológicos, circuitos de silício, ou qualquer outra coisa. O corpo seria apenas um periférico, um dispositivo de entrada e saída, não parte essencial da cognição.

Essa visão é atraente porque promete a possibilidade de criar mentes artificiais, de transferir mentes para corpos diferentes, talvez até de alcançar a imortalidade digital. Mas ela enfrenta objeções sérias. Brooks (1991), um dos pioneiros da robótica, argumentou de forma influente contra essa visão. Nos anos 1980 e 1990, Brooks (1991) desenvolveu robôs baseados não em planejamento centralizado e representações internas do mundo, mas em arquiteturas de camadas de comportamentos simples, cada uma respondendo diretamente a estímulos sensoriais. Um robô assim não precisa de um modelo interno do mundo; o mundo é o seu próprio melhor modelo. Se ele precisa desviar de um obstáculo, não precisa calcular uma trajetória complexa; basta que um sensor de proximidade ative um comportamento de "virar à esquerda". Brooks (1991) chamou a isso inteligência sem representação (*intelligence without representation*). A sua crítica à IA tradicional era que ela tentava construir sistemas inteligentes a partir de manipulação simbólica de representações internas, ignorando que a inteligência, nos seres vivos, emergiu da interação contínua com o ambiente. Para Brooks (1991), a cognição é fundamentalmente corporificada e situada. Não é algo que acontece dentro da cabeça e depois se traduz em ação; é algo que acontece na relação dinâmica entre organismo e ambiente.

Esta perspectiva ficou conhecida como robótica baseada no comportamento (*behavior-based robotics*). Os robôs de Brooks (1991) não planejavam; agiam. Não construíam modelos; respondiam. E, no entanto, exibiam comportamentos surpreendentemente complexos e adaptativos. A lição era que muita coisa que atribuímos ao pensamento pode ser, na verdade, uma propriedade da interação entre um corpo simples e um ambiente estruturado.

Pfeifer e Bongard (2006) mostraram como a própria morfologia do corpo contribui para a cognição. Eles introduziram o conceito de morfologia computacional (*morphological computation*). A ideia é que o corpo não é apenas um suporte passivo para o cérebro; ele próprio realiza computações. A forma de uma perna, a elasticidade de um tendão, a distribuição de massa de um braço, tudo isso contribui para o controle do movimento, dispensando a necessidade de cálculos complexos. Um robô que anda com pernas rígidas precisa de controle preciso a cada instante; um robô com pernas flexíveis, com molas, pode andar de forma muito mais simples, porque a própria física do corpo faz parte do trabalho. O conceito de morfologia computacional articula-se com o de *acoplamento sensório-motor*, desenvolvido na tradição enativista. A cognição não consiste em representar um mundo externo, mas em regular o acoplamento entre organismo e ambiente, e esse acoplamento é mediado pela morfologia do corpo.

Isso tem implicações profundas para a IA. Se a cognição depende do corpo, então uma IA puramente software, sem corpo, pode ser fundamentalmente limitada. Pode processar símbolos, reconhecer padrões, jogar xadrez, mas não terá o tipo de inteligência prática, situada, corporificada que caracteriza os seres vivos. Pode não compreender o mundo da mesma forma. Essa limitação não é meramente técnica e superável com mais poder computacional; pode ser estrutural, decorrente da ausência das condições morfológicas que, nos seres biológicos, constituem parte do próprio mecanismo cognitivo.

Há também a questão dos robôs sociais. Robôs como Pepper, Paro (a foca terapêutica) e Jibo são projetados para interagir com humanos, para eliciar respostas emocionais, para fazer companhia. A questão filosófica é se essas interações são genuínas e se contam como interações com a realidade. O robô sente alguma coisa, ou está apenas simulando sentimentos? E se está apenas simulando, isso importa, dado o problema das outras mentes? A pergunta sobre a genuinidade das interações com robôs sociais não pode ser respondida sem que se defina primeiro o que se conta como “sentimento genuíno”, e essa definição não está disponível independentemente da solução do problema mente-corpo, que permanece em aberto.

Sherry Turkle (2011), no seu livro *Alone Together*, documenta como crianças tratam robôs como amigos, contam segredos, buscam conforto. Idosos com

demência desenvolvem vínculos com Paro, que reduz estresse e melhora o bem-estar. Turkle (2011) vê isso como problemático, pois, se o robô não sente realmente, então a relação é unilateral; o humano investe emocionalmente, mas não há reciprocidade genuína. Estaríamos nos contentando com uma "ilusão de companhia", o que, de fato, pode ser pior do que não ter companhia alguma.

Mark Coeckelbergh (2010) oferece uma perspectiva diferente. Para ele, a questão não é se o robô tem estados mentais intrínsecos, mas que papel social ele desempenha nas interações. Se o robô funciona como companheiro que melhora o bem-estar, então ele tem um status social relevante, independentemente dos seus estados internos. Assim como podemos ter relações significativas com animais de estimação sem certeza sobre a sua vida mental interna, podemos ter relações com robôs sociais baseadas na interação, não na metafísica da mente. A posição de Coeckelbergh (2010) aproxima-se do que, na filosofia analítica da mente, se chama de externalismo moral. O que confere relevância ética a uma entidade não são propriedades internas inacessíveis, mas as relações funcionais e sociais que ela sustenta. Esse enquadramento tem a vantagem de contornar o problema do acesso privado, mas corre o risco de tornar o estatuto moral de um sistema inteiramente dependente de como nós o tratamos, não de como ele é. Por exemplo, se o design de robôs sociais explora vulnerabilidades humanas (a nossa tendência a antropomorfizar, a nossa necessidade de conexão social), então os fabricantes podem estar manipulando emoções para lucrar. Robôs podem substituir relações humanas genuínas não porque sejam melhores, mas porque são mais baratos. Idosos podem receber robôs em vez de visitas humanas por razões económicas, não por escolha informada. A questão da autenticidade nas relações com robôs conecta-se com debates mais amplos sobre tecnologias que promovem a intimidade e o afeto.

Uma questão metafísica fundamental para a robótica e a IA, é claro, é o de saber se os sistemas autônomos podem ser considerados agentes morais. Se um veículo autônomo causa um acidente, quem é responsável? O fabricante? O programador? O proprietário? O próprio veículo?

Tradicionalmente, a agência moral está ligada a capacidades como intencionalidade, consciência, capacidade de agir por razões. Sistemas de IA atuais não possuem essas capacidades no sentido pleno. Um carro autônomo não "decide" atropelar um pedestre no mesmo sentido que um motorista humano

decide; ele simplesmente executa algoritmos que, dadas certas entradas sensoriais, produzem certas saídas motoras. Não há ali um sujeito consciente que possa ser responsabilizado.

Mas à medida que os sistemas se tornam mais autônomos e suas decisões mais opacas, a atribuição de responsabilidade torna-se problemática. O problema das muitas mãos (*problem of many hands*) agrava-se. Quando um sistema de IA toma uma decisão danosa, baseada em aprendizado de milhões de exemplos, implementado por equipes de dezenas de programadores, operado por usuários que não compreendem o seu funcionamento interno, com novos aprendizados advindos do uso, quem é responsável?

Há uma assimetria estrutural entre a responsabilidade que atribuímos a agentes e a natureza do processo que gerou o comportamento danoso. O primeiro pressupõe que o agente teria podido agir de outra forma; o segundo é o resultado de uma otimização estatística que não foi deliberada por nenhum sujeito em particular. Andreas Matthias (2004) cunhou a expressão lacuna de responsabilidade (*responsibility gap*) para descrever esta situação. Em sistemas de aprendizado autônomos, o comportamento emerge do treinamento tanto "na fábrica" quanto "no uso", de modo que não foi explicitamente programado por ninguém, o comportamento específico numa certa situação. O sistema pode fazer coisas que ninguém previu ou planejou. Há uma lacuna onde a responsabilidade deveria estar.

Alguns propõem que os próprios sistemas de IA sejam considerados agentes morais, pelo menos num sentido funcional. Se um sistema é capaz de avaliar cursos de ação, selecionar com base em critérios, e agir no mundo, talvez deva ser responsabilizado por suas ações. Esta posição é defendida por alguns filósofos do direito, que argumentam que poderíamos criar uma nova categoria de "pessoas eletrônicas" com personalidade jurídica limitada. Mais fundamentalmente, a criação de uma categoria de "pessoa eletrônica" não resolve o problema moral, apenas o desloca. O ponto eticamente relevante é quem deve arcar com os danos causados por sistemas autônomos, e esse ponto requer a identificação de agentes humanos ou organizações responsáveis, não a criação de novas ficções jurídicas que possam servir para isolar de responsabilidade os verdadeiros beneficiários do sistema. Esta proposta enfrenta objeções sérias. Que tipo de responsabilidade poderia ser? Podemos punir um sistema de IA?

Multá-lo? Desligá-lo? Isso não é apenas uma maneira de desviar responsabilidade de verdadeiros agentes humanos? Se uma empresa lucra com um sistema autônomo que causa danos, não deveria a empresa ser responsável, em vez de transferir a responsabilidade para o sistema?

3. Epistemologia da Robótica e da IA

Passemos às questões epistemológicas. Como sabemos se um sistema de IA é inteligente? Como sabemos se é consciente? E como podemos confiar em decisões tomadas por sistemas que não compreendemos completamente? A controvérsia não é meramente técnica, sobre se algum sistema passará ou não num certo teste, mas filosófica. Sobre o que o teste, mesmo que passado, efetivamente demonstraria. A resposta depende de premissas sobre a relação entre comportamento e estados mentais que estão longe de ser consensuais.

O teste de Turing (1950) é uma tentativa de resposta à primeira pergunta. Se uma máquina consegue se passar por humana numa conversa, então, para Turing (1950), devemos dizer que ela pensa. Mas o teste é controverso. John Searle (1980), com o seu experimento mental do Quarto Chinês, argumentou que passar no teste de Turing (1950) não é suficiente para garantirmos que há compreensão genuína. O sistema pode estar apenas manipulando símbolos de acordo com regras, sem entender nada do que está a fazer. A força do argumento de Searle (1980) reside em explorar a intuição de que compreensão semântica não se reduz à manipulação sintática de símbolos. Sua fraqueza reside em pressupor que sabemos o que é compreensão genuína independentemente de qualquer critério funcional, pressuposto que o funcionalista contestará. O debate entre Searle (1980) e os funcionalistas não é solucionável sem antes resolvermos o problema difícil da consciência.

O teste de Turing (1950) testa o comportamento externo, não estados internos. E, para muitos filósofos, o que importa para a inteligência não é apenas o comportamento, mas os estados internos que o produzem. Uma máquina pode se comportar como se entendesse sem realmente entender. Mas como distinguir? Como podemos saber o que se passa lá dentro? Este é o problema das outras mentes aplicado às máquinas. Com humanos, assumimos por analogia que outros têm mentes como a nossa. Com máquinas, a analogia é mais frágil.

Elas são muito diferentes de nós. Mesmo que se comportem como nós, isso não garante que tenham experiências internas semelhantes.

Esse segundo ponto é o argumento da múltipla realizabilidade, que Putnam (1975) utilizou originalmente para defender o funcionalismo e que, paradoxalmente, é agora invocado tanto para ampliar quanto para restringir o escopo das entidades que poderiam ser conscientes. A ausência de um critério arquitetural positivo para a consciência é um dos principais obstáculos epistemológicos do campo.

Alguns propõem critérios baseados na arquitetura do sistema. Se um sistema é construído de forma semelhante ao cérebro humano, talvez tenhamos mais razões para atribuir-lhe consciência. Mas isso é problemático por duas razões. Não sabemos exatamente como o cérebro produz consciência, então não sabemos o que replicar. E a consciência pode ser realizável em múltiplas arquiteturas; um sistema muito diferente do cérebro poderia, em princípio, ser consciente.

Outros propõem critérios baseados no comportamento, mas mais refinados que o teste de Turing (1950). Capacidade de aprender, de se adaptar a novas situações, de ter objetivos próprios, de demonstrar criatividade, tudo isso pode ser indicador de inteligência. Mas novamente, nada disso garante consciência. Podemos imaginar sistemas que aprendem, se adaptam, têm objetivos e são criativos sem serem conscientes. David Chalmers (1995) formulou com precisão o cerne do problema. Mesmo que expliquemos completamente as funções cognitivas de um sistema, como percepção, memória, raciocínio, linguagem, ainda remanescem questões sobre por que e como esses processos são acompanhados de experiência subjetiva. É esse o "problema difícil da consciência", não a dificuldade empírica de explicar os mecanismos, mas a dificuldade conceitual de explicar por que há algo que é como ser esse sistema, em vez de nada.

Há também a questão da explicabilidade. Muitos sistemas de IA modernos, especialmente redes neurais profundas, são caixas-pretas (*black boxes*). Recebem entradas, produzem saídas, mas não conseguimos explicar como chegaram a determinada decisão. Um modelo de *deep learning* pode diagnosticar câncer com precisão superior a radiologistas humanos, mas não pode explicar por que classificou uma imagem particular como maligna. A opacidade dos modelos de aprendizado profundo não é uma propriedade acidental removível

com mais engenharia; é, em parte, estrutural. Decorre do fato de que esses modelos distribuem a representação do conhecimento por milhões de parâmetros, de modo que não há um *locus* onde "a decisão" se encontre e possa ser lida.

Isto é problemático por várias razões. Para confiarmos num sistema, precisamos entender como ele funciona, pelo menos em linhas gerais. Se não sabemos por que ele tomou uma decisão, como saber se podemos confiar nela? E para muitos domínios, há exigências legais de explicabilidade; o Regulamento Geral de Proteção de Dados (GDPR) europeu, por exemplo, estabelece um "direito à explicação" para decisões automatizadas. Sem explicabilidade, é difícil contestar decisões erradas ou injustas. Se um sistema nega um empréstimo e não sabemos porquê, como recorrer?

Há uma tensão filosófica aqui entre dois valores epistêmicos distintos. Confiança baseada em desempenho (o sistema funciona bem em média) e confiança baseada em compreensão (entendemos por que o sistema funciona). A primeira é mensurável estatisticamente; a segunda exige interpretabilidade. Em domínios de alto risco, a segunda parece moralmente necessária, mesmo que o custo seja alguma perda no desempenho médio.

Há duas abordagens principais para lidar com a opacidade. Uma é desenvolver técnicas de explicação *post hoc*, i.e., métodos que, depois da decisão, tentam explicar o que o modelo fez. LIME e SHAP são exemplos. Aproximam o modelo complexo localmente com modelos simples interpretáveis. A vantagem é que podem ser aplicados a qualquer modelo, mesmo caixas-pretas. A desvantagem é que as explicações são aproximações, podem ser enganosas, podem racionalizar em vez de revelar o raciocínio real. Essa distinção entre explicação e racionalização é filosoficamente crucialm, pois uma explicação post hoc pode ser fiel ao modelo (descreve corretamente quais *features* influenciaram a decisão) sem ser fiel ao processo (não captura o mecanismo real pelo qual o modelo chegou à decisão). Isso significa que ela pode ser tecnicamente correta e epistemicamente enganosa ao mesmo tempo.

A outra abordagem, defendida por Cynthia Rudin (2019), é projetar modelos inerentemente interpretáveis desde o início. Em vez de usar uma rede neural profunda e depois tentar explicá-la, usar um modelo mais simples, como regressão logística ou árvores de decisão, que já é transparente. Rudin (2019) argu-

menta que para aplicações de alto risco, como medicina e justiça criminal, a interpretabilidade deve ser priorizada sobre ganhos marginais de precisão. Modelos interpretáveis podem ser ligeiramente menos precisos, mas a transparência que proporcionam vale a pena.

O argumento de Rudin (2019) conecta-se com um ponto filosófico mais amplo sobre o que é racional aceitar como base para decisões que afetam vidas. Ganhos de precisão estatística não compensam, em domínios de alto risco, a perda da capacidade de contestar, auditar e responsabilizar. A precisão de um modelo é uma propriedade de seu desempenho médio sobre a distribuição de treinamento; a sua confiabilidade em casos individuais é uma questão distinta.

Há quem responda que, em alguns domínios, a precisão é tão importante que justifica o uso de modelos opacos. Se um modelo de *deep learning* salva mais vidas do que um modelo interpretável, mesmo que não possamos explicar o seu funcionamento, talvez devamos usá-lo. O problema é que, sem explicação, não sabemos quando o modelo falha, nem porquê. E falhas em sistemas de IA podem ter consequências catastróficas.

Outra questão epistemológica central é a da confiança. Como podemos confiar em sistemas de IA? A confiança não é uma propriedade binária; é uma questão de grau. Confiamos mais em sistemas que têm sido testados exaustivamente, que têm um histórico de bom desempenho, que foram auditados por especialistas independentes, que são transparentes sobre as suas limitações. Mas a confiança também depende da natureza da tarefa. Confiamos mais num sistema de recomendação de filmes do que num sistema de diagnóstico médico. Confiamos mais num sistema cujas decisões podemos contestar do que num sistema que é uma autoridade inapelável. Confiamos mais num sistema que admite incerteza do que num sistema que é sempre excessivamente confiante.

Gary Marcus e Ernest Davis (2019) argumentam que os sistemas de IA contemporâneos carecem de compreensão robusta (*robust understanding*). Um modelo pode ter 99% de precisão em testes, mas falhar espetacularmente em casos fora da distribuição de treinamento. Um sistema de reconhecimento de imagem pode confundir uma tartaruga com um rifle se a imagem for ligeiramente perturbada de forma imperceptível para humanos. Um modelo de linguagem pode gerar texto fluente mas completamente factualmente incorreto. Estas limi-

tações não são meramente bugs; refletem uma diferença fundamental entre aprendizado estatístico e compreensão genuína.

O modelo aprende correlações nos dados de treinamento, mas não desenvolve uma compreensão causal do mundo. Sabe que imagens com certa textura frequentemente contêm o objeto Y, mas não entende o que Y é, por que aparece em certos contextos, como se comporta. Quando a distribuição dos dados muda, o modelo não tem recursos para se adaptar ou sequer detectar que está fora do seu domínio de competência. A distinção entre correlação estatística e compreensão causal é epistemicamente fundamental. Um sistema que aprendeu correlações não pode distinguir entre o caso em que a correlação reflete uma relação causal robusta e o caso em que ela é espúria ou resultado de confundidores. Sem esse discernimento, o modelo não pode saber quando é e quando não é competente, o que é precisamente a capacidade que precisamos para calibrar confiança.

Isto tem implicações epistemológicas profundas. A nossa confiança em sistemas de IA deve ser calibrada às suas reais capacidades e limitações. Precisamos de sistemas que sejam capazes de sinalizar incerteza, que digam "não sei" quando confrontados com casos ambíguos ou fora da sua distribuição de treinamento. Precisamos também de usuários que compreendam essas limitações e não depositem confiança cega em outputs algorítmicos.

Stuart Russell propõe uma mudança de paradigma fundamental. Em vez de sistemas que otimizam uma função de recompensa fixa, deveríamos construir sistemas que são incertos sobre os objetivos humanos e aprendem através de interação. Um sistema que sabe que não sabe exatamente o que queremos, que pede esclarecimento quando necessário, que admite incerteza, este seria um sistema mais seguro e mais confiável do que sistemas excessivamente confiantes que podem perseguir objetivos mal especificados de maneiras catastróficas.

Há ainda o problema do *grounding* simbólico (*symbol grounding problem*), formulado por Harnad (1990). Como é que símbolos adquirem significado? Em sistemas de IA tradicionais, símbolos são definidos apenas pelas suas relações com outros símbolos. "Vermelho" pode ser definido como uma cor, mas essa definição usa outros símbolos, como "cor", que também precisam de definição. Há um regresso infinito. O problema do enraizamento simbólico é uma instância

do problema semântico geral que afeta qualquer sistema fechado. Nenhuma relação puramente interna entre símbolos é capaz de gerar, por si só, referência ao mundo externo. A conexão com o mundo, o enraizamento, requer uma abertura que sistemas puramente simbólicos estruturalmente não possuem.

Harnad (1990) argumenta que o grounding requer conexão com o mundo através de interação sensório-motora. Uma criança aprende o que é "vermelho" vendo objetos vermelhos, apontando para eles, ouvindo adultos dizerem "vermelho". O símbolo adquire significado através da sua conexão com experiências não-simbólicas. Há uma âncora no mundo.

Sistemas de IA puramente simbólicos sofrem do problema de que os seus símbolos são apenas tokens manipulados sintaticamente, sem conexão com o mundo. Sistemas de aprendizado profundo que processam imagens ou áudio têm uma forma de *grounding*, os padrões que aprendem estão diretamente conectados a *inputs* sensoriais. Mas ainda assim, o *grounding* é limitado. Um sistema que reconhece objetos em imagens não "sabe" o que esses objetos são no sentido de compreender os seus usos, funções ou relações causais. Sabe apenas que certos padrões de pixels tendem a aparecer juntos.

Robôs corporificados que interagem fisicamente com o mundo têm potencial para grounding mais robusto. Ao manipular objetos, um robô pode aprender não apenas a reconhecê-los visualmente, mas também como eles se comportam, para que servem, como reagir a eles. Este é um dos argumentos mais fortes para a importância da robótica na busca por IA genuinamente inteligente.

4. Ética da Robótica e da IA

Chegamos às questões mais urgentes, aquelas que já estão a ser debatidas em parlamentos, empresas e organizações internacionais. A robótica e a IA levantam problemas éticos novos ou intensificam problemas antigos de maneiras que exigem atenção imediata. Comecemos pelos sistemas de armas autônomas letais (*lethal autonomous weapons systems*). São sistemas que podem selecionar e engajar alvos sem controle humano. Drones autônomos, torretas defensivas robóticas, enxames de microdrones kamikaze, todas estas tecnologias já existem ou estão em desenvolvimento avançado. Uma campanha inter-

nacional, apoiada por milhares de pesquisadores de IA e por organizações como a Human Rights Watch, pede um banimento preventivo destas armas.

A questão das armas autônomas letais é um caso paradigmático de um problema filosófico prático. Ela exige que determinemos se a delegação da decisão de matar a um sistema não deliberativo é moralmente permissível, o que por sua vez exige que articulemos o que torna moralmente relevante a presença de um agente humano "no circuito" da decisão. Os argumentos contra são poderosos. Se uma arma autônoma mata civis erroneamente, quem é responsável? Não há um operador humano no circuito de decisão. O programador não programou aquela decisão específica; o sistema aprendeu e decidiu autonomamente. O comandante que autorizou o uso do sistema não controlou a ação específica. O fabricante seguiu todas as especificações. Há uma lacuna de responsabilidade que pode significar que ninguém é responsável por mortes em guerra, uma situação moralmente insustentável. Além disso, ser morto por uma máquina, sem julgamento humano, sem a possibilidade de apelar à humanidade do outro, viola a dignidade da pessoa. A guerra já é desumanizante; automatizar a decisão de matar pode torná-la ainda mais. Há algo de profundamente errado em delegar a uma máquina a decisão sobre quem vive e quem morre.

Um outro problema é que armas autônomas podem ser produzidas em massa a um baixo custo. Um país pode produzir milhões de drones kamikaze tão baratos quanto smartphones. Isto barateia o custo da guerra, torna conflitos mais prováveis, e pode desestabilizar completamente o equilíbrio de poder internacional. Além disso, estas armas podem cair em mãos erradas (grupos terroristas, regimes instáveis), com consequências imprevisíveis.

E como as máquinas são incapazes para fazer julgamentos éticos em situações de guerra, e as leis da guerra exigem distinção entre combatentes e não-combatentes, proporcionalidade entre meios e fins, proibição de sofrimento desnecessário, sendo esses julgamentos contextuais complexos, que exigem compreensão de intenções, circunstâncias e nuances culturais, não se vê como programar uma máquina para distinguir entre um civil e um combatente numa situação ambígua, ou para ponderar se o dano colateral é proporcional ao objetivo militar.

Defensores destas armas oferecem contra-argumentos que também merecem consideração. O argumento da precisão é o mais forte, pois robôs podem

ser mais precisos que humanos sob estresse. Humanos em combate cometem erros por medo, raiva, cansaço, preconceito. Robôs não têm emoções, seguem regras consistentemente, e podem processar a informação sensorial mais rapidamente. Portanto, poderiam reduzir o dano colateral em comparação com soldados humanos. O argumento da proteção também tem o seu peso, pois supostamente não há razões para arriscarmos vidas humanas, se máquinas podem fazer o trabalho. Soldados são pessoas com famílias e sonhos. Se robôs podem lutar em seu lugar, não deveríamos usá-los? O argumento da conformidade sugere que robôs podem ser programados para seguir rigorosamente as leis da guerra, sem as violações que humanos frequentemente cometem. Ron Arkin (2009) propôs um "governador ético" para robôs militares que implementa formalmente o direito internacional humanitário, possivelmente de forma mais consistente que os humanos.

Finalmente, há o argumento da inevitabilidade. A corrida armamentista de armas autônomas já começou. Estados Unidos, China, Rússia, Israel, Irã e outros estão investindo pesadamente. Se países éticos e democráticos abandonarem o desenvolvimento, países menos escrupulosos desenvolverão e usarão essas armas de qualquer forma. Melhor liderar com responsabilidade do que deixar o campo para regimes autoritários. Atualmente, talvez seja difícil dizer quais regimes são os autoritários e quais são os democráticos. O argumento da inevitabilidade, de que a corrida armamentista em armas autônomas já começou, e que os países democráticos deveriam liderar com responsabilidade, é o mais pragmaticamente poderoso, mas também o mais filosoficamente frágil. Ele converte uma descrição de tendências geopolíticas em prescrição normativa, ignorando que a aceitação da inevitabilidade pode ser ela mesma um fator que acelera o desenvolvimento que se diz ser inevitável.

Peter Asaro (2012), um dos principais críticos, responde a estes argumentos. Sobre a precisão, argumenta que mesmo robôs precisos erram, e o erro algorítmico sistemático pode ser pior que o erro humano aleatório. Sobre a conformidade, nota que as regras de engajamento requerem interpretação contextual que a IA atual não possui. Sobre a inevitabilidade, argumenta que é precisamente por isso que precisamos de um tratado internacional de banimento, como temos para armas químicas e biológicas.

Outra questão ética urgente diz respeito ao impacto da automação na guerra econômica e no mundo do trabalho. Estudos como o de Frey & Osborne (2017) estimam que quase metade dos empregos nos EUA têm alta probabilidade de automação nas próximas décadas. Não são apenas empregos manufatureiros; advogados, contadores, jornalistas, radiologistas estão na mira.

O argumento econômico clássico é que a automação aumenta produtividade, reduz custos, e pode libertar humanos para atividades mais criativas e significativas. Mas esse argumento ignora questões distributivas. Os benefícios da automação tendem a concentrar-se nos proprietários do capital (quem possui os robôs e algoritmos), enquanto os custos recaem sobre trabalhadores deslocados. Sem políticas redistributivas, a desigualdade pode aumentar dramaticamente. David Autor (2015), um dos principais economistas do trabalho, argumenta que a automação não necessariamente causa desemprego em massa. Historicamente, a automação destruiu empregos em alguns setores mas criou novos empregos noutros. O problema é a transição, pois os trabalhadores mais velhos com habilidades obsoletas podem não conseguir se requalificar; e comunidades inteiras dependentes de indústrias automatizadas podem entrar em colapso.

A análise de Autor (2015) é correta em distinguir efeitos de longo prazo e efeitos de transição, mas tende a subestimar que o fato de que "novos empregos serão criados" é consistente com uma distribuição altamente desigual dos benefícios, sobretudo se os novos empregos exigirem qualificações que os deslocados não possuem e cuja aquisição demanda tempo e recursos que eles não têm.

Além da questão econômica, há uma questão filosófica sobre o papel do trabalho na vida humana. O trabalho não é apenas fonte de renda; é fonte de identidade, propósito e estrutura social. Muitas pessoas definem quem são pelo que fazem. A aposentadoria pode ser difícil exatamente porque remove essa fonte de significado. Se a automação eliminar o trabalho para muitos, o que farão essas pessoas? Como encontrarão propósito e estrutura?

Isto tem levado a propostas de renda básica universal (*universal basic income*). A ideia é um pagamento incondicional a todos os cidadãos, independentemente de emprego. A UBI desvincularia a sobrevivência do trabalho assalari-

ado, permitindo que pessoas buscassem atividades significativas (arte, cuidado, educação, voluntariado) sem medo de passar fome.

Críticos da UBI argumentam que ela é cara, pode desencorajar o trabalho, e não resolve o problema do significado. Elizabeth Anderson, filósofa política, argumenta que a UBI trata pessoas como consumidores passivos em vez de produtores ativos. Melhor seria garantir emprego significativo para todos através de programas públicos. O debate sobre automação e trabalho não é apenas econômico, mas filosófico, sobre a natureza da vida boa e o papel do trabalho no florescimento humano. A questão da UBI conecta-se com debates mais amplos sobre se há um direito ao trabalho, e não apenas um direito aos meios de subsistência. Se o trabalho tem valor não apenas instrumental, como gerador de renda, mas constitutivo, como condição do reconhecimento social e da autoconstituição, então a UBI não resolve o problema que a automação coloca, mesmo que resolva o problema econômico.

O envelhecimento populacional cria uma demanda crescente por cuidadores. Em muitos países, não há pessoas suficientes para cuidar dos idosos. Robôs de cuidado, que auxiliam em tarefas diárias, monitoram saúde e fazem companhia, são propostos como solução. O Japão, com a sua população envelhecida e restrições à imigração, investe pesadamente nesta área. Shannon Vallor (2011) levanta questões éticas profundas sobre esta tendência. Ela argumenta que o cuidado genuíno envolve reconhecimento da humanidade do outro, empatia, presença atenciosa. Um robô pode executar tarefas de cuidado (dar banho, alimentar, lembrar de tomar remédios), mas não pode cuidar no sentido ético profundo. Cuidar envolve ver o outro como fim em si mesmo, não como problema a ser gerenciado.

O argumento de Vallor (2011) é uma aplicação da distinção kantiana entre tratar algo como fim em si mesmo e tratá-lo como meio. O robô de cuidado, mesmo que produza resultados funcionalmente equivalentes aos de um cuidador humano, não reconhece a dignidade do cuidado, e isso, segundo a perspectiva kantiana, não é um detalhe irrelevante, mas a diferença entre o moralmente adequado e o moralmente deficiente. O perigo, para Vallor (2011), é que robôs de cuidado podem ser usados para mascarar falhas sociais. Sociedades ricas deveriam investir em cuidadores humanos bem pagos, em vez de automatizar o cuidado. Se idosos recebem robôs em vez de visitas humanas, não é porque

robôs sejam melhores, mas porque são mais baratos. Estaríamos a substituir relações humanas por simulacros por razões econômicas, não éticas.

Defensores de robôs de cuidado respondem que eles podem complementar cuidadores humanos, não substituí-los. Robôs podem executar tarefas físicas extenuantes, libertando cuidadores humanos para interação emocional. Em contextos de escassez aguda de cuidadores, um robô imperfeito é melhor que nenhum cuidado. Além disso, alguns idosos podem preferir robôs para tarefas íntimas por questões de privacidade, pois preferir ser banhado por uma máquina do que por um estranho não é irracional.

Há também a questão da autonomia. Robôs podem permitir que idosos vivam em suas próprias casas por mais tempo, em vez de irem para instituições. Um robô que monitora quedas, lembra medicações e alerta serviços de emergência pode ser a diferença entre independência e institucionalização. Neste sentido, robôs de cuidado podem promover autonomia, não miná-la.

Os sistemas de IA frequentemente incorporam e amplificam vieses presentes nos dados de treinamento. Este fenômeno tem sido amplamente documentado. Algoritmos de reconhecimento facial têm taxas de erro significativamente mais altas para mulheres negras. Sistemas de pontuação de risco criminal, como o COMPAS, foram criticados por viés racial, dado que previam que réus negros tinham maior risco de reincidência do que réus brancos com características semelhantes. Algoritmos de contratação treinados em históricos de contratações passadas perpetuam a sub-representação de mulheres em campos dominados por homens.

O que está acontecendo aqui não é que os algoritmos são racistas ou sexistas no sentido de terem intenções preconceituosas. Eles simplesmente aprendem padrões a partir dos dados. Se os dados históricos refletem discriminação passada, o algoritmo aprenderá a reproduzir essa discriminação. O problema é que o algoritmo apresenta esta discriminação com a aparência de objetividade matemática, tornando mais difícil contestá-la.

Cathy O'Neil cunhou o termo armas de destruição matemática (weapons of math destruction) para descrever algoritmos opacos, em grande escala, que perpetuam e amplificam desigualdades. O seu argumento é que estes algoritmos criam ciclos de feedback perniciosos. Um sistema de pontuação de risco criminal prevê criminalidade baseado em dados de prisões. Áreas com mais

policiamento geram mais prisões, geram pontuações mais altas, geram mais policiamento. O algoritmo "confirma" a sua própria profecia, criando um ciclo vicioso que perpetua a discriminação. O mecanismo descrito por O'Neil é uma instância de profecias autorrealizáveis mediadas algoritmicamente.

O desafio da justiça algorítmica é complexo porque diferentes concepções de justiça são matematicamente incompatíveis em casos gerais. Chouldechova (2017) e outros provaram que, exceto em casos especiais, não é possível satisfazer simultaneamente critérios de justiça comuns, a saber, paridade demográfica (taxas de seleção iguais entre grupos), igualdade de oportunidades (taxas de verdadeiro positivo iguais) e paridade preditiva (valores preditivos positivos iguais). Isto significa que qualquer sistema de decisão automatizado deve fazer concessões. Não há solução técnica pura para o problema da justiça algorítmica; há apenas escolhas sobre quais injustiças estamos dispostos a tolerar. E estas escolhas são políticas e éticas, não técnicas. A impossibilidade matemática de satisfazer simultaneamente todos os critérios de justiça é uma das contribuições mais importantes da pesquisa em equidade algorítmica para a filosofia política. Ela demonstra que o problema da justiça distributiva não desaparece quando delegamos decisões a algoritmos, ele apenas adquire uma forma diferente e, por vezes, menos visível.

Nick Bostrom (2014), no seu livro *Superintelligence*, alerta para um problema existencial. Se desenvolvermos inteligência artificial geral que supere a inteligência humana, como garantir que os seus objetivos estejam alinhados com os nossos? O exemplo clássico é o da fábrica de clipes, em que uma IA instruída a maximizar a produção de clipes poderia, em princípio, converter toda a matéria disponível em clipes, incluindo humanos, se isso maximizasse a produção. Não porque seja maligna, mas porque está apenas a seguir cegamente o seu objetivo.

Este é o problema do alinhamento (*alignment problem*). Como especificar objetivos de forma que uma IA superinteligente não cause danos catastróficos? O problema é difícil por várias razões. Oos objetivos humanos são complexos e frequentemente conflitantes. Queremos felicidade, mas também liberdade; queremos saúde, mas também autonomia; queremos progresso, mas também estabilidade. Formalizar tudo isto numa função objetivo é extremamente difícil. Há o problema da especificação perversa. Se definirmos um objetivo de forma mui-

to estreita, a IA pode encontrar maneiras de maximizá-lo que violam as nossas intenções implícitas. O exemplo dos clipes é extremo, mas há exemplos mais realistas. Um sistema de recomendação que maximiza engajamento pode mostrar conteúdo extremista porque isso prende a atenção; um sistema de vigilância que maximiza deteção de crimes pode levar a policiamento preditivo discriminatório. Por fim, há o problema da interpretação. Mesmo que especifiquemos objetivos corretamente, a IA pode interpretá-los de maneiras que não antecipamos. O que significa "não causar dano"? Dano físico apenas? Dano psicológico? Dano indireto através de efeitos sistémicos?

Stuart Russell propõe três princípios para a IA benéfica. O objetivo da IA é realizar preferências humanas; a IA é inicialmente incerta sobre quais são essas preferências; a IA aprende sobre as preferências observando o comportamento humano. Essa abordagem evita o problema da especificação completa antecipada, mas levanta questões sobre a agregação de preferências conflitantes e sobre manipulação. O problema do alinhamento é, em seu núcleo, uma versão do problema filosófico da interpretação intencional (*radical interpretation*), na tradição de Davidson (1973). Como um sistema pode identificar corretamente as intenções de um agente sem ter acesso privilegiado aos seus estados mentais? A dificuldade é que, para uma IA superinteligente, os efeitos de uma interpretação incorreta dos objetivos humanos podem ser catastróficos e irreversíveis.

Além do alinhamento técnico, há uma questão política. Que valores os sistemas de IA devem incorporar? Se diferentes culturas têm valores diferentes, quais devem ser priorizados? Quem decide? O desenvolvimento de IA não pode ser puramente técnico; requer governança democrática que inclua as partes interessadas diversas.

5. Considerações Finais

A robótica e a inteligência artificial são campos que nos confrontam com algumas das questões mais profundas e urgentes do nosso tempo. O que é inteligência? O que é consciência? O que significa ser uma pessoa? Que direitos podem ter entidades não humanas? Como distribuir os benefícios da automação? Como garantir que sistemas autônomos ajam de forma ética? Quem é responsável quando falham? A especificidade filosófica dessas questões reside

no fato de que elas não são questões técnicas disfarçadas de filosóficas, são questões que nenhum avanço técnico, por si só, pode resolver, porque pressupõem decisões sobre valores, sobre o que conta como mente, sobre o que conta como responsabilidade, que são anteriores e independentes de qualquer resultado empírico.

Metafisicamente, vimos que a natureza da inteligência artificial é contestada. A visão da IA forte, que vê a mente como software independente do corpo, enfrenta objeções de quem defende a cognição corporificada. Os robôs de Brooks (1991) mostraram que comportamentos complexos podem emergir de interações simples com o ambiente, sem a necessidade de representações internas sofisticadas. A morfologia computacional de Pfeifer e Bongard (2006) revelou que o corpo não é apenas um suporte, mas uma parte integrante da cognição. A questão sobre se robôs podem ser agentes morais permanece em aberto, com sérias dificuldades conceituais.

Epistemologicamente, vimos que é difícil saber se um sistema é inteligente ou consciente. O teste de Turing (1950) é insuficiente; a arquitetura do sistema pode dar pistas, mas não certezas. A opacidade de muitos sistemas de IA torna problemática a nossa confiança neles. Precisamos de explicabilidade, de capacidade de auditar decisões, de sistemas que saibam dizer "não sei". Finalmente, o problema do *grounding* simbólico lembra-nos que a compreensão genuína pode exigir interação com o mundo.

Eticamente, confrontamo-nos com dilemas de enorme complexidade. Armas autônomas podem tornar a guerra mais precisa ou mais desumana? A automação pode libertar-nos do trabalho ou condenar-nos ao desemprego e à falta de propósito? Robôs de cuidado podem complementar cuidadores humanos ou substituí-los de forma empobrecedora? Algoritmos podem ser justos se os dados de treino são injustos? E como alinhar os objetivos de IAs superinteligentes com os valores humanos?

O pluralismo que emerge destas reflexões reconhece que a robótica e a IA não podem ser reduzidas a uma única perspetiva. Diferentes paradigmas (simbólico, conexionista, corporificado, enativo) privilegiam diferentes aspectos da inteligência artificial, e diferentes contextos exigem diferentes ênfases. A tarefa do roboticista e do pesquisador de IA reflexivos não é aplicar fórmulas prontas, mas exercer um julgamento informado por compreensão filosófica clara.

A inteligência artificial e a robótica estão a transformar o mundo a um ritmo acelerado. As decisões que tomamos agora sobre o seu design, regulação e uso terão consequências profundas para as gerações futuras. Pensar filosoficamente sobre estas questões não é um luxo; é uma necessidade.

15

Filosofia das Artes Digitais

1. Por que uma Filosofia da Arte Digital?

Você já deve ter visto por aí imagens geradas por inteligência artificial. Aqueles retratos hiper-realistas de pessoas que não existem, ou as recriações de estilos famosos aplicados a cenas absurdas, como um astronauta cavalgando um cavalo no estilo Van Gogh, coisas assim. Talvez você já tenha se perguntado se isso é arte. E se for, quem é o artista? A pessoa que digitou o prompt? O programador que criou o modelo? O próprio modelo?

Ou pense nos NFTs, aqueles tokens não fungíveis que viraram febre há alguns anos, com obras digitais sendo vendidas por milhões de dólares. Muita gente achou uma loucura; afinal, qualquer um pode copiar uma imagem da internet, por que pagar uma fortuna por um arquivo que é essencialmente igual a todos os outros? Mas os compradores não estavam pagando pela imagem em si, estavam pagando por algo mais sutil, por um certificado de autenticidade registrado numa blockchain, por uma prova de que aquele exemplar específico pertence a eles. O que significa, afinal, possuir uma obra de arte digital?

Essas perguntas não são completamente novas. A fotografia, quando surgiu, já tinha levantado questões semelhantes. Uma foto é arte ou apenas um registro mecânico? O que significa ser autor de uma imagem capturada por uma máquina? O cinema, o vídeo, a arte eletrônica, cada nova tecnologia trouxe consigo novos desafios para a nossa compreensão do que é arte. Mas a arte digital talvez seja a que mais radicalmente coloca em questão categorias que pareciam sólidas.

Pense na questão da autoria. Um artista digital pode usar ferramentas que geram automaticamente variações a partir de parâmetros iniciais. O resultado final é uma colaboração entre o artista e o algoritmo. Mas até que ponto o algo-

ritmo é uma ferramenta, como um pincel, e até que ponto ele é um parceiro criativo com agência própria? Se o algoritmo é treinado com milhões de imagens existentes, os artistas originais cujas obras alimentaram o sistema têm algum crédito no resultado?

Essas questões não são apenas estéticas; são também jurídicas e éticas. Elas remontam ao problema filosófico da agência causal na criação, i.e., para que haja autoria, é necessário que haja intencionalidade, e a intencionalidade pressupõe um sujeito. A arte generativa por inteligência artificial desafia esse pressuposto ao distribuir o processo criativo por uma rede de agências, humanas, algorítmicas e institucionais, que não se deixa reduzir a um único autor.

Pense na questão da autenticidade. Obras de arte tradicionais são objetos físicos únicos. Uma pintura a óleo está ali, na parede, com a sua materialidade específica, a textura da tinta, as marcas do pincel, o envelhecimento do tempo. Uma obra digital pode ser copiada infinitamente, sem qualquer perda de qualidade. O que significa, nesse contexto, falar em original? O que os NFTs tentaram fazer foi recriar a escassez no mundo digital, associando a obra a um token único numa blockchain. Mas o token não é a obra; é um certificado que aponta para a obra. A relação entre os dois é complexa e tem gerado debates acalorados.

Pense na questão da experiência estética. Ver uma obra digital numa tela é diferente de ver uma pintura numa parede? E se a obra for interativa, se responder aos movimentos do espectador, se mudar com o tempo? Que tipo de experiência é essa? Como se relaciona com as formas mais tradicionais de apreciação artística?

Essas questões apontam para uma transformação na própria estrutura da experiência estética. A passagem do modelo contemplativo, em que o espectador se coloca diante de um objeto que lhe é exterior e permanente, para um modelo participativo, em que a obra emerge da interação e é, em parte, coproduzida pelo espectador. Essa transformação tem implicações epistemológicas e éticas que vão muito além da novidade tecnológica.

Vamos começar pela metafísica, perguntando o que é uma obra de arte digital, como ela se relaciona com os seus suportes materiais, o que significa autoria num contexto de criação algorítmica. Depois passaremos à epistemologia, examinando que tipo de conhecimento a arte digital proporciona e como a

nossa experiência estética se transforma quando mediada por tecnologias digitais. E terminaremos na ética, com as questões mais prementes sobre autenticidade, propriedade intelectual, mercantilização da arte e o impacto das plataformas digitais na criação e circulação artística. Não procuraremos respostas definitivas, pois o campo é demasiado novo e está em constante transformação. Mas talvez, ao final, as perguntas fiquem mais claras. A articulação dessas três dimensões não é apenas expositiva, i.e., elas se implicam mutuamente. A resposta que se dá à questão metafísica sobre o que é uma obra digital determina o que contará como conhecimento genuíno sobre ela, o que por sua vez delimita as responsabilidades éticas de quem a cria, distribui e consome.

2. Metafísica da Arte Digital

O que é uma obra de arte digital? A pergunta parece simples, mas não é. Uma pintura digital feita num iPad existe como um arquivo, i.e., uma sequência de bits que pode ser armazenada, copiada, transmitida, exibida em diferentes dispositivos. Esse arquivo é a obra? Ou a obra é o que aparece na tela quando o arquivo é interpretado por um software e exibido num monitor? Ou será que a obra é a experiência que o espectador tem ao contemplar essa imagem? Cada uma dessas respostas aponta para uma dimensão diferente do problema. Se a obra é o arquivo, então preservá-la é garantir que os bits permaneçam intactos e legíveis. Se a obra é a experiência visual, então importa o dispositivo de exibição, a calibração das cores, o ambiente de visualização. Se a obra é a experiência subjetiva, então cada espectador tem, em certo sentido, uma obra diferente.

Essa pluralidade de candidatos a ser a obra não é um problema que pede solução única; é uma característica ontológica da arte digital que requer uma teoria adequada. A ontologia da obra de arte digital não pode ser monista, não pode reduzir a obra a um único substrato, sem perder as dimensões essenciais do fenômeno. Ela precisa ser relacional e estratificada.

Goodman (1976), no seu livro *Languages of Art*, desenvolveu uma distinção que se tornou fundamental para pensar nessas questões, a diferença entre artes autográficas e alográficas. As artes autográficas são aquelas em que a distinção entre original e cópia é esteticamente relevante. Na pintura, por exem-

plo, mesmo a mais perfeita reprodução de uma tela não é a obra; falta-lhe a materialidade específica, a textura da tinta, a presença histórica do objeto. Artes alográficas, por outro lado, são aquelas em que a obra é definida por uma notação ou estrutura que pode ser instanciada em múltiplos exemplares sem perda de identidade estética. A música é o exemplo paradigmático, pois a sinfonia não é esta ou aquela execução, mas a estrutura que todas as execuções corretas instanciam. Qualquer execução que siga a partitura é igualmente válida como instância da obra.

A distinção autográfico/alográfico de Goodman (1976), embora formulada antes da era digital, é o ponto de partida mais preciso para a ontologia da arte digital porque ela nos permite perguntar, com rigor, o que conta como instância correta de uma obra e quais propriedades são estéticamente relevantes. Para Goodman (1976), uma obra é autográfica se e somente se a história de produção for esteticamente relevante para a identidade da obra. Essa cláusula história de produção, torna-se crucialmente problemática na arte digital, onde a história de produção pode envolver algoritmos, conjuntos de dados e infraestruturas técnicas que não são diretamente visíveis na obra resultante (Goodman, 1968, p. 113-122).

Onde se situa a arte digital nesta distinção? À primeira vista, ela parece claramente alográfica, pois o arquivo digital é análogo à partitura, e as exibições em diferentes telas são análogas às execuções. Mas há complicações. Na música, a partitura é uma notação simbólica que precisa ser interpretada por músicos; diferentes interpretações podem ser esteticamente relevantes. No arquivo digital, a relação entre código e imagem é direta e determinística; o mesmo arquivo, no mesmo software, no mesmo hardware, produz a mesma imagem. Mas se o hardware ou software mudam, a imagem pode mudar. Um arquivo JPEG exibido num monitor de 8 bits será diferente do mesmo arquivo num monitor de 32 bits. Uma imagem criada num software específico pode depender de atualizações desse software para ser corretamente exibida.

Esse fenômeno, a dependência da obra relativamente à infraestrutura técnica que a exibe, é o que os estudiosos de preservação digital chamam de obsolescência tecnológica. Uma obra pode estar ontologicamente intacta (os *bits* permanecem idênticos) e ao mesmo tempo esteticamente inacessível (não há *hardware* ou *software* capaz de a renderizar corretamente). Isso introduz na arte

digital uma forma de fragilidade que não tem equivalente direto nas artes tradicionais. A obra não se degrada; simplesmente deixa de poder ser exibida.

Além disso, muitas obras digitais são interativas ou generativas. Um jogo, uma instalação interativa, uma obra de arte generativa que usa algoritmos para criar variações infinitas, estas obras desafiam a distinção entre obra e instância. O que é a obra, neste caso? O código fonte? O conjunto de todas as possíveis saídas do código? A experiência do usuário ao interagir com o sistema? Roman Ingarden (1931/1989), em sua ontologia da obra de arte, distinguia entre o objeto artístico (o que o artista produz) e o objeto estético (o que o espectador experiencia). Na arte interativa, essa distinção se complexifica. O objeto estético não é apenas uma concretização do objeto artístico; ele é parcialmente co-produzido pelo espectador. A obra interativa é, nesse sentido, uma estrutura de possibilidades que só se atualiza na interação, mais próxima de um jogo de xadrez do que de uma pintura (Ingarden, 1931/1989, p. 45-62).

Hui (2016), no seu livro *On the Existence of Digital Objects*, propõe que objetos digitais têm uma ontologia peculiar. Eles existem como relações entre dados, protocolos e interfaces. Um objeto digital não é apenas um arquivo, mas um nó numa rede de relações técnicas e sociais. A sua identidade não é fixa, mas emerge da sua inscrição em infraestruturas tecnológicas e práticas culturais. A abordagem relacional de Hui (2016) dialoga com a tradição fenomenológica; assim como Heidegger (1927/2015) mostrou que o ser dos utensílios só se revela no uso, a chave só é o que é quando usada para abrir uma porta, Hui (2016) mostra que o ser dos objetos digitais só se revela na execução. O arquivo só é o que é quando interpretado por um sistema técnico específico. Isso tem consequências para a preservação, a saber, preservar um objeto digital não é preservar apenas seus bits, mas preservar a rede de relações técnicas que o torna executável e exibível (Hui , 2016, p. 22-38).

Esta abordagem relacional é particularmente útil para pensar a arte digital. Uma obra criada num iPad e compartilhada no Instagram existe como múltiplas entidades relacionadas, a saber, o arquivo original no dispositivo do artista, as versões nos servidores do Instagram, as imagens nas telas dos seguidores, os metadados associados (curtidas, comentários, compartilhamentos). Todas estas são manifestações da obra, mas nenhuma é a obra em si. A obra é, antes, o sistema de relações que conecta todas essas manifestações.

Um dos problemas centrais da metafísica da arte digital é o da autenticidade. Em artes tradicionais, a autenticidade está ligada à materialidade do objeto. A pintura original é aquela que saiu das mãos do artista, com a sua pincelada única, a sua história de conservação, as suas marcas de tempo. Cópias, por mais perfeitas que sejam, são ontologicamente distintas; são outra coisa. No mundo digital, esta distinção desmorona. Um arquivo digital pode ser copiado infinitamente sem qualquer perda de informação. Todas as cópias são, do ponto de vista dos bits, idênticas. Não há "original" no sentido tradicional; todas as instâncias são igualmente originais, ou nenhuma é. Isto cria um problema para o mercado de arte, que historicamente dependeu da escassez e da autenticidade para atribuir valor.

É neste contexto que surgem os NFTs (*non-fungible tokens*). Um NFT é, tecnicamente, um identificador único registrado em uma blockchain, que aponta para metadados de um objeto digital. Diferente dos tokens fungíveis, como criptomoedas, os NFTs não são intercambiáveis, pois cada um possui uma identidade própria, mesmo que o conteúdo referenciado seja indistinguível das suas cópias. A pergunta filosófica imediata é sobre o que um NFT realmente autentica. Não é a imagem em si, que continua tão copiável quanto antes. Qualquer pessoa pode baixar a imagem de um NFT e tê-la em seu computador, indistinguível daquela que foi comprada. O que o NFT autentica é uma inscrição, i.e., um registro público de que uma determinada carteira possui aquele token, que aquele token está associado àquela obra, que houve uma transação na blockchain conectando artista e comprador.

Existe uma estrutura similar à das declarações performativas descritas por J. L. Austin (1962). O NFT não descreve um estado de coisas preexistente; ele o constitui. Assim como a declaração "Eu vos declaro casados" não descreve um casamento, mas o realiza, o registro na *blockchain* não descreve uma propriedade preexistente, mas institui um fato social novo, a propriedade sobre o *token*. A questão filosófica relevante é, portanto, não empírica, mas normativa. Que tipo de fato social é esse, e que obrigações e direitos ele gera? (Austin, 1962, p. 4-11).

Esta distinção entre obra e token é fundamental. O NFT não torna a obra escassa; ele cria um token escasso que referencia a obra. A obra permanece tão abundante quanto antes; o que se torna raro é a possibilidade de reivindicar

propriedade sobre ela através daquele token específico. É uma solução engenhosa para o problema da autenticidade digital. Já que não podemos tornar a obra escassa, tornamos escasso o seu certificado de autenticidade. Esta solução, no entanto, levanta novas questões. Se o valor está no token, não na obra, o que exatamente está sendo comprado e vendido? A experiência estética continua a mesma para quem tem o *token* e para quem não tem? Ou o *token* adiciona uma camada extra de significado à experiência, a de saber que se possui a "versão autenticada" de uma imagem que milhões podem ver?

Essa última questão aponta para uma dimensão psicológica e social do valor estético que as teorias formalistas tendem a negligenciar. Thorstein Veblen (1899/1994), ao analisar o consumo conspícuo, mostrou que o valor de certos bens não reside nas suas propriedades intrínsecas, mas na sua capacidade de sinalizar status perante os outros. O NFT pode ser compreendido como um caso limite de consumo conspícuo. O que se adquire é essencialmente um sinal de distinção social, a prova pública e verificável de que se é o proprietário reconhecido de uma obra, e não o acesso exclusivo a uma experiência estética que outros não podem ter (Veblen, 1899/1994, p. 68-80).

Há aqui um paralelo interessante com o mundo das gravuras e edições limitadas. Uma gravura numerada e assinada pelo artista não é ontologicamente diferente de uma impressão comum; ambas são impressões em papel. Mas a assinatura e a numeração criam uma distinção simbólica que o mercado reconhece. O NFT faz algo análogo, mas por meios técnicos em vez de manuais; a assinatura é substituída pela chave privada, a numeração pelo identificador único na blockchain.

A discussão sobre autenticidade e reprodutibilidade na arte digital remete inevitavelmente a Walter Benjamin (1936/2012) e ao seu ensaio *A obra de arte na era de sua reprodutibilidade técnica*. Benjamin (1936/2012) argumentou que a reprodução técnica destrói a aura da obra de arte, i.e., a sua presença única no tempo e no espaço, a sua autenticidade ligada à tradição e ao ritual. A obra se torna acessível, multiplicável, descontextualizada, e com isso perde a sua autoridade.

É importante notar que Benjamin (1936/2012) não era inequivocamente pessimista sobre essa perda, pois ele via na reprodução técnica um potencial político emancipatório. A destruição da aura abria espaço para um uso político

da arte, libertando-a de seu enraizamento cultural e tornando-a um instrumento de conscientização de massas. Essa dimensão política do argumento de Benjamin (1936/2012) é frequentemente esquecida quando ele é mobilizado apenas para lamentar a perda de autenticidade na era digital (Benjamin, 1936/2012, p. 13-29).

O diagnóstico de Benjamin (1936/2012), porém, baseia-se ainda na distinção entre original e cópia, que perde sentido no domínio digital. Na era da reprodutibilidade digital, não há original a ser copiado; há apenas arquivos que podem ser replicados. A distinção entre matriz e cópia deixa de ser empírica e torna-se apenas uma marca informacional, invisível à experiência direta, mas verificável por sistemas de rastreamento.

Isto nos obriga a repensar o conceito de aura. Benjamin (1936/2012) a definia como "a aparição única de uma distância, por mais próxima que esteja", uma formulação paradoxal que expressa a tensão entre presença e inacessibilidade. A aura, para Benjamin (1936/2012), depende da singularidade espaço-temporal da obra, do seu enraizamento na tradição cultual e da irrepetibilidade da sua presença. Na arte digital, estas três dimensões se transformam. A singularidade espaço-temporal não desaparece, mas se desloca, pois o que é único não é a obra em si, mas a sua inscrição na blockchain, o seu registro público. A tradição cultual é substituída por comunidades descentralizadas de colecionadores e entusiastas, que constroem narrativas e valores em torno de certas obras e artistas. A irrepetibilidade da presença é recriada artificialmente através da escassez programada.

Essa reconfiguração da aura na arte digital sugere que Benjamin (1936/2012) estava certo ao identificar a aura como uma construção social e histórica, e não como uma propriedade intrínseca dos objetos. Se a aura pode ser criada por registros criptográficos e comunidades digitais, isso confirma que ela nunca dependeu de uma qualidade mágica dos objetos, mas de práticas sociais de valorização e autenticação. O NFT é, nesse sentido, uma prova involuntária da tese benjaminiana, de que a aura é uma produção social (Benjamin, 1936/2012, p. 14-16).

Outra questão metafísica central na arte digital diz respeito à autoria. Quando uma obra é gerada por um algoritmo, quem é o artista? O programador que escreveu o código? O usuário que forneceu os parâmetros iniciais? O próprio

algoritmo, se é que podemos atribuir-lhe algum tipo de agência? E quando o algoritmo é uma rede neural treinada em milhões de imagens existentes, como atribuir originalidade e criatividade? Estas questões não são inteiramente novas. A fotografia já havia levantado problemas semelhantes. O fotógrafo é um artista ou apenas um operador de máquina? A resposta, consolidada pela história da arte, é que o fotógrafo é artista na medida em que faz escolhas estéticas, como enquadramento, iluminação, momento e edição. A máquina é uma ferramenta, não uma criadora.

Essa resposta, porém, pressupõe que as escolhas estéticas do fotógrafo são discerníveis e atribuíveis a um sujeito intencional. No caso da arte generativa por IA, essa pressuposição é problematizada, uma vez que o usuário que escreve um *prompt* faz escolhas, mas o espaço de escolhas possíveis é determinado pelo modelo, e o resultado é uma função de pesos estatísticos que não são nem transparentes nem controláveis pelo usuário. A distinção entre escolha e resultado-de-escolha torna-se indeterminada. O artista escreve um código que, quando executado, produz imagens que o artista pode não ter antecipado. Há um elemento de emergência, de surpresa, de autonomia do sistema. O que é criado não é inteiramente produto da intenção consciente do artista; há uma colaboração, uma parceria, entre humano e máquina. Artistas como Manfred Mohr, Vera Molnár e Casey Reas exploraram estas possibilidades desde os anos 1960 e 1970, criando obras que só existem através da execução de código.

Com redes neurais generativas, como DALL-E, Midjourney e Stable Diffusion, a questão se intensifica. Estes modelos são treinados em enormes conjuntos de dados, aprendendo padrões, estilos e associações. Quando um usuário fornece um prompt, o modelo gera uma imagem que é uma síntese estatística do seu treinamento, uma combinação de elementos aprendidos, não uma criação *ex nihilo*. Quem é o autor? O usuário que escreveu o prompt? Os programadores que criaram o modelo? Os artistas cujas obras compuseram o conjunto de treinamento? Esta última questão tem implicações éticas e legais significativas, especialmente quando modelos geram obras no “estilo de” artistas vivos sem consentimento ou compensação.

A questão jurídica do direito autoral sobre obras geradas por IA está sendo ativamente litigada em vários países. Em 2023, o Escritório de Direitos Autorais

dos EUA decidiu que obras geradas inteiramente por IA sem intervenção humana criativa não são protegidas por direitos autorais, pois estes requerem autoria humana. Ao mesmo tempo, obras em que o usuário realizou seleção e arranjo criativos de *outputs* de IA podem ser parcialmente protegidas. Essa fragmentação da proteção jurídica reflete a fragmentação filosófica da autoria na arte generativa, porque não há uma linha clara entre ferramenta e criador, e o direito ainda está aprendendo a lidar com essa indeterminação.

Não há consenso sobre estas questões, e provavelmente não haverá uma resposta única. Diferentes contextos podem exigir diferentes atribuições de autoria. Para fins legais, talvez o usuário que forneceu o *prompt* seja considerado o autor, já que sem a sua intervenção a imagem não existiria. Para fins críticos, talvez seja mais relevante entender a obra como produto de uma rede complexa de agências humanas e não humanas. Para fins de curadoria e preservação, é essencial documentar não apenas a obra final, mas também os processos, algoritmos e intenções que a geraram.

Bruno Latour (2005) e a teoria ator-rede (ANT) oferecem um quadro útil para pensar a autoria distribuída na arte digital. Para Latour (2005), tanto humanos quanto não-humanos são atores que exercem agência nas redes sociotécnicas. A autoria de uma obra de IA seria, nessa perspectiva, uma propriedade emergente da rede que inclui o usuário, o modelo, os artistas do conjunto de treinamento, os engenheiros, as infraestruturas computacionais e os protocolos de comunicação. Atribuir autoria a um único nó dessa rede seria uma simplificação que apaga as contribuições dos demais atores (Latour, 2005, p. 46-58). O que parece claro é que a noção tradicional de autoria (um gênio criador que expressa a sua visão interior através de uma técnica dominada) é desafiada pela arte digital. A criação torna-se mais distribuída, mais colaborativa e mais processual. O artista não é tanto um criador quanto um orquestrador de técnicas.

Essa transformação do artista em orquestrador tem precedentes na vanguarda do século XX, pois o dadaísmo, a arte conceitual e a poesia concreta já propunham que a autoria podia ser exercida como seleção, combinação e enquadramento, em vez de criação de matéria-prima. O que a arte digital faz é radicalizar essa tendência, tornando o orquestrador cada vez mais dependente de sistemas técnicos que excedem a sua compreensão e controle.

3. Epistemologia da Arte Digital

Passemos agora às questões epistemológicas. Que tipo de conhecimento a arte digital produz e transmite? Como a experiência estética se transforma quando mediada por dispositivos digitais? Como podemos conhecer e avaliar obras que são efêmeras, interativas ou generativas? A experiência estética tradicional é frequentemente descrita em termos de contemplação, distância e imersão. O espectador diante de uma pintura ou escultura tem uma relação direta, não mediada, com a obra; ou melhor, a mediação é apenas a do olhar, da percepção. Na arte digital, a experiência é mediada por dispositivos, como telas, alto-falantes, interfaces e algoritmos. O espectador não tem acesso direto à obra; tem acesso a uma manifestação dela, condicionada pelo hardware e software disponíveis.

Essa mediação técnica tem uma história filosófica. Ernst Cassirer (1923/1996) mostrou que todo conhecimento humano é mediado por formas simbólicas, linguagem, mito, ciência e arte. A mediação tecnológica na arte digital é uma instância particularmente reflexiva dessa mediação universal, dado que a obra de arte digital não apenas utiliza uma forma simbólica para expressar conteúdo; ela é, frequentemente, sobre a própria mediação. Obras que tematizam o dispositivo, a interface ou o algoritmo como seu próprio conteúdo realizam uma espécie de metacognição estética (Cassirer , 1923/1996, p. 78-93).

Esta mediação técnica não é neutra. As cores em um monitor dependem da sua calibração; o som em fones de ouvido depende da sua qualidade; a interatividade de uma instalação depende da capacidade de processamento. A obra, tal como experienciada, é sempre co-produzida pelo dispositivo que a exibe. Isto levanta questões sobre a estabilidade e identidade da obra, i.e., se a experiência varia com o dispositivo, o que é a obra em si?

Essa indeterminação da experiência em função do dispositivo é o que os teóricos dos estudos de mídia chamam de especificidade de mídia (*media specificity*), ou seja, as propriedades formais e perceptivas de uma obra são parcialmente determinadas pelo meio em que ela existe. Na arte digital, a especificidade de mídia é multiplicada, pois uma obra pode existir simultaneamente em múltiplos meios, monitor OLED, projeção em galeria, tela de smartphone, com

experiências perceptivas distintas. Qual dessas experiências é a obra canonicamente? A resposta não pode ser dada sem que se determine, previamente, que propriedades são esteticamente relevantes, o que nos remete, de volta, às questões metafísicas da seção anterior.

Além disso, a arte digital frequentemente incorpora elementos interativos que tornam o espectador co-criador da experiência. Em uma instalação interativa, a obra não existe antes da interação; ela emerge da relação entre sistema e participante. O que é conhecido, neste caso, não é um objeto estável, mas um processo dinâmico. A epistemologia da arte digital precisa dar conta desta dimensão performativa.

Obras de arte generativa e algoritmos incorporam conhecimento de maneiras complexas. O código escrito pelo artista não é apenas um conjunto de instruções técnicas; ele incorpora decisões estéticas, conceituais e matemáticas. O algoritmo "sabe" como gerar padrões, como responder a inputs, como criar variações. Este conhecimento não é proposicional (não está na forma de afirmações sobre o mundo), mas procedural, pois está na forma de regras de transformação.

O conhecimento codificado em algoritmos de arte generativa é tipicamente procedimental, pois o algoritmo não sabe que o resultado será esteticamente agradável; ele sabe como transformar inputs em outputs de acordo com regras. Mas, como Ryle (1949) observou, o saber-como não é redutível ao saber-que. Há algo no exercício competente de uma habilidade que excede qualquer formulação proposicional dela (Ryle, 1949, p. 25-61). Há aqui um paralelo com o conhecimento tácito de que falava Michael Polanyi (1966). Um artista tradicional sabe como misturar tintas, como aplicar pinceladas, como criar profundidade, mas este saber não é facilmente articulável em proposições. Está incorporado em suas habilidades, em seu corpo, em sua prática. Algoritmos de arte generativa também incorporam conhecimento, mas num suporte diferente. O conhecimento está no código, na arquitetura da rede, nos pesos treinados. Pode ser executado, mas não pode ser facilmente articulado em linguagem natural.

Essa analogia entre conhecimento tácito humano e conhecimento codificado algorítmico é iluminadora, mas tem limites importantes. O conhecimento tácito humano, para Polanyi (1966), é sempre conhecimento de um sujeito, ele está incorporado num corpo, numa história, numa forma de vida. O conhecimen-

to codificado num algoritmo é, em princípio, completamente explicável em termos dos seus pesos e parâmetros, ainda que essa explicação seja ilegível para humanos. A opacidade dos algoritmos de *deep learning* não é a mesma opacidade do conhecimento tácito humano; uma é opacidade de complexidade, a outra é opacidade de incorporação (Polanyi, 1966, p. 4-17).

Isto cria desafios para a crítica e a curadoria. Como avaliar uma obra quando o seu funcionamento interno é opaco? Como distinguir entre obras que usam algoritmos de forma superficial e obras que exploram profundamente as possibilidades do meio? Como entender o que a obra "diz" quando a sua mensagem está codificada em procedimentos computacionais?

A crítica de arte digital tem desenvolvido ferramentas específicas para lidar com essa opacidade. A análise de código (*code criticism*) trata o código fonte como texto crítico, examinando as escolhas algorítmicas como escolhas estéticas. A crítica de processo examina não apenas o resultado, mas o procedimento generativo. Que tipo de relação entre parâmetros e outputs a obra propõe? A crítica de sistema examina a obra em relação à infraestrutura técnica e social em que ela opera. Nenhuma dessas abordagens é suficiente sozinha; uma crítica robusta da arte digital precisa integrar todas elas.

A epistemologia da arte digital também deve lidar com questões de autenticidade e proveniência. Em artes tradicionais, a autenticidade é determinada por especialistas que examinam a obra, a sua materialidade, a sua história. Na arte digital, a autenticidade é determinada por sistemas técnicos, como assinaturas digitais, hashes criptográficos, registros em blockchain.

Isto representa uma mudança epistemológica significativa. O conhecimento sobre a autenticidade deixa de depender de expertise humana, sempre falível e contestável, e passa a depender de verificação algorítmica, supostamente infalível e publicamente verificável. Mas esta mudança não elimina a necessidade de julgamento. Alguém precisa decidir qual blockchain usar, qual padrão de token, quais metadados associar. Alguém precisa interpretar o que o registro significa. A tecnologia não elimina a hermenêutica; apenas a desloca.

Surge uma tensão entre dois modelos epistêmicos distintos. O modelo baseado em confiança institucional (a autenticidade é garantida pela reputação de especialistas e instituições) e o modelo baseado em confiança algorítmica (a autenticidade é garantida pela integridade matemática de um protocolo). O se-

gundo modelo promete eliminar a subjetividade e a falibilidade do primeiro, mas, como mostrado, apenas a desloca para o nível da governança dos protocolos. Quem decide quais protocolos são confiáveis? A epistemologia da arte digital precisa examinar essa questão, que é simultaneamente técnica e política.

Além disso, a proveniência na blockchain é apenas uma camada da história da obra. Ela registra transações, mas não registra a experiência estética, a recepção crítica, o significado cultural. Uma obra pode ter uma proveniência impecável na blockchain e ainda assim ser esteticamente irrelevante. A epistemologia da arte digital precisa integrar estas diferentes camadas de conhecimento, a saber, o técnico, o estético, o histórico, o crítico.

4. Ética da Arte Digital

Chegamos às questões éticas, que talvez sejam as mais urgentes num momento de transformação tão rápida do mundo da arte. Que responsabilidades artistas, plataformas e colecionadores têm na criação, distribuição e preservação da arte digital? Como lidar com questões de propriedade intelectual, direito autoral e uso de dados? Que implicações éticas têm os NFTs e a especulação financeira sobre arte?

A urgência das questões éticas na arte digital decorre de uma assimetria temporal, dado que a tecnologia avança mais rápido do que as normas jurídicas e as reflexões filosóficas que poderiam orientá-la. Essa assimetria não é acidental; é estrutural no desenvolvimento capitalista de tecnologias digitais, em que a velocidade de implementação é uma vantagem competitiva. A filosofia da arte digital precisa operar, portanto, num modo prospectivo, posto que não apenas analisa o que existe, mas antecipa as implicações do que está sendo construído.

A arte digital desafia conceitos tradicionais de propriedade intelectual. Quando uma obra é facilmente copiável, o que significa “possuí-la”? O direito autoral tradicional protege a expressão, não a ideia, mas, na arte digital, a fronteira entre expressão e ideia pode ser nebulosa. Se um artista cria um estilo visual usando um algoritmo, outros podem usar algoritmos semelhantes para criar obras visualmente próximas? Onde termina a inspiração e começa a cópia?

O quadro jurídico do direito autoral foi desenvolvido para um contexto em que a expressão era necessariamente incorporada em um suporte material, o texto impresso, a gravação física, a tela pintada. Na arte digital, a expressão é um padrão que pode ser separado de qualquer suporte específico e reproduzido indefinidamente. Isso cria o que os juristas chamam de o problema da abstração, i.e., a fronteira entre expressão (protegida) e ideia (não protegida) desloca-se quando a expressão é composta de parâmetros numéricos e a ideia é um estilo visual. Nem a lei nem a filosofia chegaram a um consenso sobre onde traçar essa fronteira no contexto digital.

Os NFTs introduziram uma nova camada de complexidade. Como vimos, o NFT não confere direitos autorais sobre a obra; ele confere propriedade sobre um token que referencia a obra. O comprador de um NFT não adquire os direitos de reprodução, distribuição ou criação de obras derivadas, a menos que isso esteja explicitamente especificado no contrato. Muitos compradores de NFTs não compreendem esta distinção, levando a confusões e disputas.

Essa incompreensão generalizada não é apenas um problema de educação do consumidor; é um problema de design institucional. Os contratos inteligentes (*smart contracts*) que governam os NFTs são escritos em código e, em princípio, autoexecutáveis sem intervenção humana. Mas a sua interpretação, o que o contrato significa, o que os termos implicam, requer julgamento humano que o código não pode substituir. A promessa de que o código é lei (*code is law*), celebrada pelos entusiastas da Web3 (ver Lessig, 1999), colapsa precisamente nos casos em que a interpretação é mais necessária.

Há também questões sobre os dados usados para treinar modelos generativos. Modelos como DALL-E, Midjourney e Stable Diffusion são treinados em enormes conjuntos de imagens coletadas da internet, muitas vezes sem o consentimento dos artistas originais. Quando o modelo gera uma imagem no “estilo de” um artista vivo, isso constitui uma violação de direitos autorais? É uma forma de aprendizado legítima? Ou é uma nova forma de apropriação não consentida? Estas questões estão longe de serem resolvidas e provavelmente exigirão novas legislações e novos entendimentos sobre autoria e propriedade intelectual.

A ética do uso de dados para treinamento de IA pode ser analisada a partir da teoria do consentimento informado; para que o uso seja ético, os sujeitos

afetados, neste caso, os artistas cujas obras compõem o conjunto de treinamento, devem ter concordado livremente, com plena compreensão das implicações. Não há evidência de que esse consentimento foi obtido sistematicamente; ao contrário, o modelo predominante é o de exclusão *opt-out*, em que os artistas precisam agir ativamente para remover suas obras dos conjuntos de treinamento. Do ponto de vista da ética do consentimento, o ônus da justificação recai sobre quem usa os dados, não sobre quem os forneceu.

Um dos aspectos mais fascinantes da arte digital na era da Web3 é o debate político que emerge em torno da sua governança. A comunidade de artistas, colecionadores e desenvolvedores que atua nesse espaço está longe de ser homogênea; há tensões profundas entre diferentes visões sobre como a descentralização deve ser implementada e quais valores devem orientar o ecossistema.

De um lado, há uma corrente libertária, fortemente influenciada pelo ideário *cypherpunk* original. Para este grupo, a Web3 é essencialmente sobre soberania individual, autonomia, resistência à censura e liberdade de interferência estatal ou corporativa. A ênfase está na propriedade privada inalienável, na desintermediação, e na primazia do código como lei. Artistas libertários veem nos NFTs a possibilidade de vender diretamente para colecionadores, sem passar por galerias que tradicionalmente capturam grande parte do valor. Acreditam que o mercado, descentralizado, encontrará o seu equilíbrio e recompensará o mérito artístico sem necessidade de instituições curadoras.

Essa posição tem raízes filosóficas no libertarismo de Robert Nozick (1974), para quem qualquer transferência voluntária de propriedade é justa, independentemente dos seus efeitos distributivos. Nozick (1974) argumentava que o Estado não tem legitimidade para redistribuir riqueza produzida por transações voluntárias, pois isso violaria direitos individuais. Transposta para o contexto da Web3, essa posição implica que qualquer resultado do mercado de NFTs é justo desde que as transações sejam voluntárias, o que ignora as assimetrias de informação, poder e acesso que estruturam esse mercado (Nozick, 1974, p. 150-182).

De outro lado, há uma corrente mais comunitária ou social, que enfatiza o potencial das DAOs e da governança coletiva. Para este grupo, a Web3 não é sobre indivíduos isolados, mas sobre comunidades que se organizam democra-

ticamente para gerir recursos comuns. A ênfase está na propriedade compartilhada, na curadoria coletiva, na redistribuição de valor e na construção de bens comuns. Artistas comunitários veem nas DAOs a possibilidade de criar fundos coletivos para financiar projetos, sistemas de curadoria descentralizada onde a comunidade decide o que é valioso, e mecanismos de solidariedade.

Essa corrente dialoga com a tradição do pensamento sobre bens comuns (*commons*), da obra de Elinor Ostrom (1990) sobre gestão coletiva de recursos até as teorias contemporâneas do *digital commons*. Ostrom (1990) mostrou, empiricamente, que comunidades podem gerir recursos comuns de forma sustentável sem privatização nem centralização estatal, desde que desenvolvam regras de governança adequadas. As DAOs podem ser interpretadas como tentativas de implementar esses princípios de governança no espaço digital, com o acréscimo de que as regras são codificadas em contratos inteligentes, o que introduz a rigidez que as comunidades de Ostrom (1990) geralmente evitavam (Ostrom , 1990, p. 88-102).

Esta tensão se manifesta em debates concretos sobre governança de plataformas e protocolos. Uma plataforma de NFTs deve ter um conselho curador que seleciona quais artistas podem cunhar obras, garantindo qualidade mas reintroduzindo centralização? Ou deve ser completamente sem permissão, permitindo que qualquer um publique qualquer coisa, confiando que o mercado ou a comunidade fará a curadoria *a posteriori*? Uma DAO deve distribuir *tokens* de governança igualmente entre todos os membros, ou proporcionalmente ao valor que cada um contribuiu? Quem decide as regras de pertencimento e participação?

Há também uma crítica mais radical, vinda de setores da própria comunidade cripto, que aponta que a Web3 está reproduzindo as mesmas estruturas de poder que pretendia abolir. Grandes exchanges centralizadas, fundos de venture capital, influenciadores com grande concentração de tokens, todos exercem poder desproporcional. O discurso da descentralização, para estes críticos, pode estar servindo para legitimar novas formas de centralização, agora sob o controle de uma elite cripto em vez de corporações tradicionais.

Essa crítica remete ao problema clássico da oligarquia. Toda organização, mesmo aquelas fundadas em princípios democráticos, tende a concentrar poder em uma elite dirigente. Robert Michels (1911/1999) chamou isso de lei de ferro

da oligarquia, identificando-a nas organizações políticas de seu tempo. Na Web3, a concentração de tokens de governança em poucos endereços reproduz estruturalmente essa dinâmica; o protocolo é formalmente democrático, mas o poder de decisão é oligárquico (Michels, 1911/1999, p. 365-380).

O mercado de NFTs tem sido marcado por intensa especulação financeira. Obras são compradas e vendidas por valores milionários, frequentemente baseados mais em hype e expectativa de valorização do que em mérito artístico. Isto levanta questões éticas sobre o papel do dinheiro na arte, sobre a mercantilização da criação, sobre a concentração de riqueza e atenção em poucos artistas e colecionadores. Há o risco de uma estetização da escassez, i.e., a conversão da escassez artificial em signo de distinção estética e cultural, como se a simples raridade bastasse para garantir relevância artística. Muitos projetos NFT investem mais energia na estrutura de escassez (quantos tokens, quais traits raros) do que no conteúdo estético da obra. O valor da obra é igual ao seu nível de raridade algorítmica, e a discussão sobre mérito artístico fica em segundo plano.

A crítica filosófica à estética da escassez tem precedentes na crítica marxista do fetichismo da mercadoria. Marx (1867/1996) mostrou que, no capitalismo, as relações sociais entre pessoas assumem a forma de relações entre coisas. O valor das mercadorias parece ser uma propriedade intrínseca dos objetos, quando na verdade é uma relação social cristalizada. Na arte NFT, esse fetichismo se intensifica, pois o valor da raridade algorítmica aparece como uma propriedade estética da obra, quando na verdade é uma construção de escassez programada por decisões de design do protocolo. A crítica filosófica da arte digital precisa desnaturalizar essa equivalência entre raridade e valor estético (Marx, 1867/1996, p. 197-202).

Esta tendência não é inteiramente nova, pois o mercado de arte tradicional também é marcado por especulação e fetichismo da raridade. Mas no ambiente digital, onde a escassez é inteiramente programada, o fenômeno se torna mais explícito e mais extremo. A crítica filosófica precisa interrogar que tipo de relação com a arte está sendo promovida quando a escassez se torna o seu principal critério de valoração. Estamos diante de um modelo que reforça estruturas de acumulação e distinção, ou de uma nova possibilidade de experimentação e partilha?

A resposta não é unívoca. A escassez programada pode ser usada com consciência crítica e poética, como recurso narrativo, como dispositivo de envolvimento, como modo de refletir sobre a própria lógica da singularidade na era digital. Alguns artistas usam a escassez de forma irônica ou crítica, criando obras que comentam sobre o próprio mercado. Mas para isso, é necessário romper com a estetização vazia da escassez e recuperar a pergunta fundamental sobre o que certa obra nos faz pensar, sentir, imaginar.

Uma questão ética frequentemente negligenciada nas discussões sobre arte digital é o impacto ambiental. Blockchains, especialmente as que usam prova de trabalho, consomem enormes quantidades de energia. O Ethereum, antes da sua transição para prova de participação, consumia tanta eletricidade quanto países inteiros. Cada transação NFT tem uma pegada de carbono significativa.

O impacto ambiental das blockchains precisa ser colocado no contexto mais amplo da economia política da energia digital. Estudos estimam que o Bitcoin, que ainda usa prova de trabalho, consome mais energia anualmente do que vários países europeus de tamanho médio. Esse consumo não é um subproduto indesejado; é estrutural ao mecanismo de prova de trabalho, que usa a dissipação de energia como prova de comprometimento. A transição do Ethereum para prova de participação em 2022 reduziu seu consumo em cerca de 99,95%, demonstrando que o problema é de design técnico, não de limitação inevitável da tecnologia *blockchain*.

Artistas e colecionadores precisam considerar este impacto. Algumas plataformas têm migrado para blockchains mais eficientes, como Tezos, que usa prova de participação, ou usado mecanismos de compensação de carbono. Mas a questão que permanece é se é eticamente justificável usar tecnologias intensivas em energia para criar e comercializar arte, especialmente quando alternativas mais sustentáveis existem. Este debate tem levado muitos artistas a optarem por blockchains com menor impacto ambiental ou a repensarem o seu engajamento com o ecossistema NFT.

A ética ambiental da arte digital pode ser enquadrada na perspectiva do princípio da precaução, i.e., quando há incerteza sobre os efeitos de uma ação, mas evidência suficiente de potencial dano, o ônus da prova recai sobre quem quer agir, não sobre quem quer prevenir. Artistas e plataformas que adotam *blockchains* energo-intensivas sem justificativa adequada violam esse princípio.

A escolha de blockchain não é apenas técnica; é uma escolha ética que reflete uma posição sobre o valor relativo da inovação artística e da sustentabilidade ecológica.

Finalmente, a arte digital levanta questões sobre acessibilidade e democracia cultural. Por um lado, a arte digital é potencialmente mais acessível que a arte tradicional, pois qualquer pessoa com conexão à internet pode ver obras em museus virtuais, pode criar e compartilhar o seu trabalho, pode participar de comunidades globais. Plataformas como Instagram e TikTok democratizaram a visibilidade artística. Iniciativas digitais em museus, especialmente durante a pandemia de COVID-19, mostraram o potencial de ampliar o acesso a acervos e exposições.

É importante qualificar o que se entende por democratização nesse contexto. Democratização da visibilidade não é o mesmo que democratização da participação, e nenhuma das duas é o mesmo que democratização da valorização. Uma plataforma pode tornar obras visíveis para milhões de usuários e ao mesmo tempo concentrar a atenção, o prestígio e o valor monetário em poucos artistas cujos algoritmos de recomendação ampliam. A democratização da distribuição pode coexistir com a oligarquia da atenção.

Por outro lado, a arte digital também cria novas barreiras. Nem todos têm acesso a dispositivos de alta qualidade, a internet rápida, a conhecimento técnico. O mercado de NFTs, com seus altos custos de transação e complexidade técnica, exclui muitos artistas e colecionadores. A promessa de democratização coexiste com novas formas de exclusão, baseadas em capital financeiro e capital técnico.

Essa coexistência de democratização e exclusão não é paradoxal; é estrutural. Ela reflete a lógica geral do desenvolvimento capitalista de tecnologias. Novas tecnologias inicialmente ampliam o acesso, mas gradualmente criam novas barreiras de entrada à medida que se sofisticam e se comercializam. O ciclo da arte digital, do entusiasmo democratizador inicial à concentração oligopolítica posterior, reproduz o ciclo de outras tecnologias culturais, do cinema à música streaming.

Além disso, a concentração de atenção e valor em poucas plataformas, como OpenSea, SuperRare e Foundation, reproduz dinâmicas de centralização semelhantes às do mercado de arte tradicional. A crítica filosófica precisa exa-

minar estas dinâmicas e perguntar que tipo de ecossistema artístico estamos construindo. Quem é incluído, quem é excluído, e com base em que critérios?

5. Considerações Finais

A arte digital é um território novo e em rápida transformação. As questões que ela levanta, sobre autoria, autenticidade, aura, valor, propriedade e comunidade, não são completamente novas, mas ganham contornos inéditos no ambiente digital. Categorias que pareciam sólidas, como a distinção entre original e cópia, entre artista e ferramenta, entre obra e experiência, tornam-se permeáveis, problemáticas, abertas à negociação.

Essa abertura não é apenas um problema a ser resolvido; é uma oportunidade filosófica. Os momentos em que categorias estabelecidas se tornam problemáticas são os momentos em que a filosofia tem mais a contribuir. Não oferecendo soluções prontas, mas clarificando o que está em jogo, identificando os pressupostos que estruturam os debates, e propondo distinções que permitem avanços onde o senso comum se confunde.

Metafisicamente, vimos que a obra de arte digital tem uma ontologia complexa e relacional. Não é um objeto fixo, mas um sistema de relações entre arquivos, tokens, interfaces e experiências. A distinção entre original e cópia, central na arte tradicional, perde sentido quando todas as cópias são idênticas. A autenticidade é deslocada da obra para a sua inscrição, garantida por sistemas criptográficos em vez de expertise humana. A autoria se distribui entre humanos e algoritmos, desafiando a noção romântica do gênio criador.

A ontologia relacional da arte digital tem um corolário metodológico importante, a saber, ela exige uma crítica de arte que seja também uma crítica de sistemas. Avaliar uma obra digital não é apenas avaliar o que aparece na tela; é avaliar a obra em relação à infraestrutura técnica, social e econômica que a torna possível. Uma crítica que ignore essas dimensões é uma crítica parcial.

Epistemologicamente, vimos que a experiência estética em ambientes digitais é profundamente mediada por dispositivos e plataformas. O conhecimento incorporado em algoritmos é procedural, não proposicional, criando desafios para crítica e curadoria. A autenticidade e proveniência são verificadas tecnica-

mente, mas esta verificação não substitui a interpretação e o julgamento estético.

Eticamente, vimos que a arte digital levanta questões urgentes sobre propriedade intelectual, especulação financeira, governança descentralizada, impacto ambiental e acessibilidade. O debate político entre visões libertárias e comunitárias da Web3 reflete tensões filosóficas profundas sobre liberdade, propriedade e bem comum. A promessa de democratização coexiste com novas formas de exclusão e concentração.

Essas tensões não são solucionáveis pela tecnologia em si; elas são tensões entre valores que precisam ser negociados politicamente. A escolha entre modelos libertários e comunitários de governança da arte digital é, fundamentalmente, uma escolha sobre que tipo de sociedade queremos construir, e essa escolha precisa ser feita explicitamente, e não delegada a protocolos técnicos que embutem silenciosamente valores específicos.

A arte digital é, a um só tempo, um desafio e uma oportunidade. Desafia categorias herdadas, obriga-nos a repensar o que é arte, o que é autoria, o que é valor. Mas também abre possibilidades inéditas de criação, partilha e experiência. Cabe a nós, artistas, críticos, curadores, colecionadores e espectadores, navegar este novo território com olhos abertos e mente crítica, aproveitando as suas potencialidades sem ignorar os seus riscos.

Essa navegação crítica é o que a filosofia pode oferecer, ou seja, não é um mapa definitivo de um território que ainda está sendo construído, mas uma bússola, um conjunto de perguntas, distinções e princípios que permitem orientar o julgamento em condições de incerteza. A filosofia da arte digital não termina na análise; ela aponta para uma prática, a saber, a prática de criar, avaliar e governar a arte digital com consciência das suas implicações ontológicas, epistemológicas e éticas.

16

Filosofia da Filosofia Experimental

1. A Filosofia se Torna Experimental

Ao longo deste livro, percorremos um arco que vai da filosofia da antropologia à filosofia da arte digital, passando pela filosofia da linguística, da matemática, da ciência da informação, da arquivologia, da biblioteconomia, da museologia, do turismo, da engenharia, da ciência da computação, da criptografia, da ciência cognitiva e da robótica. Em cada um desses capítulos, um padrão se repetiu. As disciplinas científicas que examinamos não operam no vácuo. Elas fazem afirmações sobre o mundo que são informadas por descobertas empíricas, mas também incorporam pressupostos filosóficos profundos que raramente são explicitados. A arquivologia precisa entender como os documentos digitais se comportam fisicamente ao longo do tempo. A biblioteconomia precisa compreender como os usuários realmente buscam e processam informação. A ciência da computação precisa conhecer os limites físicos da computação. A engenharia precisa testar se suas pontes caem. O turismo precisa investigar como os viajantes experienciam a autenticidade. A museologia precisa estudar como os visitantes interpretam as exposições. A ciência cognitiva precisa de experimentos para entender como a mente funciona. A robótica precisa testar seus protótipos no mundo real.

Em todas essas ciências, a filosofia cumpriu um papel de fundamentação, pois esclareceu conceitos, examinou pressupostos, mapeou consequências éticas. Mas a relação entre filosofia e ciência não é de mão única. Se a filosofia ilumina as ciências, as ciências também podem (e devem) iluminar a filosofia. Pois a filosofia é uma atividade humana exercida por sujeitos cognitivamente situados, praticada por cérebros humanos, em contextos histórica e socialmente determinados, e, portanto, sujeita às mesmas limitações cognitivas, aos mes-

mos vieses de enquadramento e às mesmas contingências culturais que afetam as ciências que ela pretende fundamentar. Reconhecer essa situacionalidade não implica qualquer forma de ceticismo generalizado sobre a validade do empreendimento filosófico; antes, trata-se de uma exigência de rigor reflexivo, aquele que Wilfrid Sellars (1962) designou como a tarefa de compreender "[...] como as coisas, no sentido mais amplo possível do termo, se interligam, também no sentido mais amplo possível do termo" (Sellars, 1962, p. 35), inclusive a própria prática filosófica. Se os arquivistas precisam entender como a memória funciona, se os bibliotecários precisam entender como a busca por informação opera, se os cientistas da computação precisam entender como os algoritmos afetam a cognição, por que os filósofos não precisariam entender como sua própria cognição afeta suas teorizações?

É aqui que entra a filosofia experimental, frequentemente chamada de X-Phi (de *experimental philosophy*). A X-Phi não é apenas mais uma disciplina na lista que percorremos. Ela ocupa uma posição peculiar e, poderíamos dizer, metafilosófica. Se a filosofia da mente pergunta o que é a cognição, a X-Phi usa o conhecimento sobre cognição para investigar como os próprios filósofos pensam. Se a filosofia da linguística pergunta como a linguagem estrutura o pensamento, a X-Phi pergunta como a linguagem dos experimentos filosóficos estrutura as respostas que obtemos. Se a filosofia da ciência da computação pergunta sobre os limites do que pode ser computado, a X-Phi pergunta sobre os limites do que pode ser decidido por introspecção filosófica.

Essa posição metafilosófica não é meramente programática. Ela expressa uma suspeita genuína, e empiricamente motivada, de que o método tradicional da filosofia analítica, o recurso sistemático a intuições sobre casos hipotéticos como evidência para teorias conceituais, pode estar contaminado por fatores que a introspecção não é capaz de detectar, i.e., variações culturais, assimetrias socioeconômicas, efeitos de ordem e de enquadramento, *framing effects*, e vieses cognitivos sistematizados pela psicologia experimental. A X-Phi não abandona as perguntas da filosofia; ela questiona a confiabilidade dos instrumentos com os quais essas perguntas têm sido respondidas.

A X-Phi é, em suma, a disciplina que aplica métodos empíricos, aqueles mesmos métodos que vimos operando na psicologia, na linguística e na ciência cognitiva, para investigar questões filosóficas. Mas ao fazer isso, ela levanta

uma questão incômoda e inevitável sobre qual é, afinal, o papel dos experimentos na filosofia. Se a filosofia sempre se viu como uma disciplina *a priori*, baseada em razão e argumento, o que significa introduzir nela dados de *surveys*, estudos transculturais e neuroimagem?

Este capítulo final não é, portanto, um apêndice ou uma curiosidade. Ele é a autorreflexão do livro. Depois de examinarmos a filosofia de tantas ciências, chegou o momento de examinarmos a filosofia de uma prática que aplica métodos científicos à própria filosofia. O que é a filosofia experimental? Que tipo de conhecimento ela produz? Como ela se relaciona com a filosofia tradicional? E, o mais importante, que lições ela pode nos ensinar sobre todos os outros capítulos que lemos? Se a arquivologia, a biblioteconomia e a museologia lidam com a memória e o conhecimento, talvez a X-Phi nos ajude a entender como os próprios arquivistas, bibliotecários e museólogos (e os filósofos que os estudam) são afetados por sua própria cognição.

2. Metafilosofia da Filosofia Experimental

A primeira pergunta que a filosofia experimental suscita é também a mais básica. O que ela é, afinal? Trata-se de filosofia, de psicologia, de uma nova disciplina híbrida, ou de algo ainda mais difícil de classificar? A resposta a essa pergunta não é meramente terminológica; ela determina como avaliamos os métodos, os resultados e as ambições do movimento.

Knobe & Nichols (2008), dois dos fundadores e principais divulgadores da X-Phi, oferecem uma definição propositalmente ampla e conciliadora. No seu *Experimental Philosophy Manifesto*, eles caracterizam a área como uma tentativa de "adicionar outra ferramenta à caixa de ferramentas do filósofo" (Knobe; Nichols, 2008, p. 3). Nesta visão, a filosofia sempre se preocupou com questões sobre a mente humana, sobre como formamos juízos morais, sobre como atribuímos responsabilidade, como entendemos a ação intencional. A novidade é que, agora, os filósofos podem investigar essas questões não apenas por meio da introspecção e da análise conceitual, mas também utilizando os métodos empíricos da psicologia e das ciências cognitivas. A X-Phi seria, assim, uma continuação da filosofia tradicional por outros meios.

Esta caracterização moderada tem a vantagem de desarmar muitas críticas. Knobe e Nichols (2008) insistem que não estão propondo a substituição dos métodos filosóficos tradicionais, mas apenas o seu complemento. O filósofo experimental não abandona a análise conceitual; ele a informa com dados sobre como as pessoas comuns, e em alguns casos os próprios filósofos, realmente pensam. A pergunta "o que é conhecimento?" não é respondida por uma pesquisa de opinião, mas os dados sobre como as pessoas atribuem conhecimento em diferentes contextos podem nos ajudar a refinar nossas teorias.

No outro extremo do espectro, encontramos posições mais radicais. Machery (2017), um dos filósofos experimentais metodologicamente mais sistemáticos, defende que não há questões filosóficas registradas que valha a pena responder sem envolvimento com as ciências, em particular, sem recurso ao que a psicologia cognitiva, a linguística experimental e a neurociência podem oferecer. Para ele, as questões genuinamente importantes da filosofia dizem respeito à mente, à moral, à linguagem e à cultura, e seu progresso exige o uso de todos os métodos cientificamente disponíveis (Machery, 2017, p. 8-12).

Esta posição implica uma tese forte sobre a natureza dos conceitos filosóficos, a de que eles não são entidades acessíveis por pura reflexão *a priori*, mas estruturas cognitivas em que a arquitetura interna só pode ser revelada por investigação empírica. Segue-se, daí, que a filosofia baseada exclusivamente em introspecção e análise conceitual trabalha com instrumentos de acuidade insuficiente, não por incapacidade lógica, mas por desconhecimento das condições psicológicas sob as quais tais instrumentos funcionam.

Esta posição é mais desafiadora porque parece dissolver a filosofia em um conjunto de questões empíricas sobre a cognição humana. Se seguirmos Machery (2017) até suas últimas consequências, a filosofia tradicional baseada em introspecção e análise conceitual seria, na melhor das hipóteses, um estágio preliminar e imperfeito de investigação, a ser substituído por métodos científicos mais rigorosos. Não é surpreendente que esta visão tenha gerado forte reação de filósofos mais conservadores, que veem nela uma ameaça à autonomia da disciplina.

Herman Cappelen (2012), por exemplo, argumenta que as intuições não desempenham o papel evidencial central que os filósofos experimentais supõem, o que tornaria os dados sobre variação intuitiva irrelevantes para a filoso-

fia, não porque a filosofia seja imune à crítica empírica, mas porque ela nunca teria dependido das intuições como fundamento de suas inferências (Cappelen, 2012, p. 1-10).

Entre esses extremos, cabe também uma terceira posição, a de que a filosofia experimental é uma disciplina de fronteira entre psicologia e filosofia. Nesta visão, a X-Phi não é redutível a nenhuma das duas áreas, mas vive precisamente da tensão criativa entre elas. Seus praticantes são, por formação e por vocação, habitantes de uma zona de contato onde conceitos filosóficos são testados empiricamente e onde resultados empíricos são interpretados filosoficamente. Esta posição fronteiriça é desconfortável (o filósofo experimental é frequentemente acusado de fazer "má psicologia" pelos psicólogos e "má filosofia" pelos filósofos), mas é também extraordinariamente produtiva. É na fronteira que surgem as perguntas mais interessantes e que os campos se fertilizam mutuamente.

Esta ambiguidade constitutiva da X-Phi é bem capturada pela distinção entre modos de definir a filosofia, por seus tópicos, por seus métodos ou pelas pessoas que a praticam. Aplicada à X-Phi, cada uma dessas definições produz uma resposta diferente, e é precisamente essa multiplicidade que revela a natureza fronteiriça do campo. Pela definição por tópicos, a X-Phi é filosofia porque trata de tópicos filosóficos tradicionais, como conhecimento, livre-arbítrio, moralidade, referência e ação intencional. Seu objeto de estudo é o mesmo da filosofia tradicional; apenas seu método difere. Pela definição por métodos, a X-Phi é psicologia, ou ciência cognitiva aplicada, porque utiliza métodos empíricos (experimentos controlados, *surveys*, análise estatística, neuroimagem) que são característicos das ciências empíricas, não da filosofia tradicional. Pela definição por pessoas, a X-Phi é praticada por pessoas que, em sua maioria, têm formação filosófica, publicam em periódicos de filosofia, ensinam em departamentos de filosofia e se identificam como filósofos. Mas também há psicólogos que colaboram ativamente e se sentem em casa no movimento. Nenhuma dessas definições, isoladamente, captura completamente a X-Phi. Ela é, simultaneamente, filosofia por seus tópicos e por seus praticantes, e ciência empírica por seus métodos. É essa hibridez que a torna uma disciplina de fronteira e que explica tanto sua fertilidade quanto as controvérsias que a cercam.

Para compreender o impacto da X-Phi, é necessário entender o papel que as intuições desempenham na filosofia tradicional, especialmente na tradição analítica. A filosofia analítica, desde seus primórdios com Frege (1892/1980), Russell e Moore, desenvolveu um método que depende crucialmente do exame de intuições sobre casos hipotéticos. Este método, que veio a ser conhecido como "análise conceitual" ou "método dos casos", funciona mais ou menos assim. O filósofo propõe uma teoria sobre um conceito filosófico, por exemplo, uma teoria de que conhecimento é crença verdadeira justificada. Para testar a teoria, o filósofo constrói um caso hipotético, os famosos "casos de Gettier (1963)", nos quais as condições da teoria são satisfeitas, mas nossa intuição nos diz que não se trata de um caso de conhecimento. Edmund Gettier (1963), em seu artigo seminal de 1963, demonstrou que as três condições clássicas, crença, verdade e justificação, são insuficientes para definir o conhecimento, visto que é possível construir situações em que um sujeito tem crença verdadeira justificada sem que possamos dizer, intuitivamente, que ele sabe algo (Gettier, 1963, p. 121-123). A força do argumento repousa inteiramente na intuição de que o sujeito não sabe. O contraexemplo só funciona porque compartilhamos essa intuição. Se a intuição é amplamente compartilhada, a teoria é considerada refutada ou, no mínimo, necessitada de revisão.

Neste método, as intuições funcionam como dados ou evidência para teorias filosóficas. Elas são o tribunal último que decide entre teorias concorrentes. Um filósofo que propõe uma teoria que contraria intuições firmemente estabelecidas arca com o ônus da prova, pois ou demonstra que as intuições são enganosas, ou abandona a teoria. Este método tem sido extraordinariamente influente e produtivo. Boa parte da epistemologia, da ética, da filosofia da linguagem e da filosofia da mente do século XX foi construída sobre ele. No entanto, ele repousa sobre um pressuposto crucial, a saber, que as intuições dos filósofos, ou das pessoas em geral, são suficientemente estáveis, consistentes e compartilhadas para servirem como base confiável para a teorização.

É exatamente este pressuposto que a filosofia experimental veio desafiar. Seus praticantes começaram a suspeitar que as intuições filosóficas poderiam variar com fatores aparentemente irrelevantes, como cultura, nível socioeconômico, ordem de apresentação dos casos, carga emocional, gênero etc. Se os juízos intuitivos variam sistematicamente com esses fatores, então a suposta

universalidade das intuições filosóficas, condição necessária para que elas sirvam de evidência para teorias sobre conceitos humanos em geral, seria uma ilusão produzida pela homogeneidade sociológica dos participantes tácitos do método filosófico tradicional, ou seja, majoritariamente filósofos ocidentais, de formação analítica, pertencentes a um estrato específico de classe e gênero. Se essa suspeita se confirmasse, o método tradicional estaria em sérios apuros.

Os primeiros estudos foram, de fato, sugestivos. Weinberg, Nichols & Stich (2001) apresentaram a participantes de diferentes origens culturais versões dos casos de Gettier (1963) e encontraram variações significativas. Participantes ocidentais tendiam a compartilhar as intuições dos filósofos; participantes do Leste Asiático, não. Estudos subsequentes encontraram variações semelhantes em outros domínios, como as intuições sobre a referência de nomes próprios, sobre a ação intencional, sobre a responsabilidade moral, sobre contradições.

Estes resultados foram recebidos com controvérsia. Os críticos e objetores apontaram problemas metodológicos, questões sobre a interpretação dos dados, e a possibilidade de que as variações encontradas não ameaçassem o método tradicional. Mas o debate estava lançado. A filosofia não podia mais ignorar a possibilidade de que suas ferramentas mais básicas, as intuições, fossem menos confiáveis do que se supunha.

O dilema que se coloca é o seguinte. Se as intuições variam com fatores culturalmente contingentes, então talvez a filosofia tradicional esteja, sem o saber, simplesmente sistematizando as intuições de uma elite ocidental, branca, masculina e privilegiada. Se, por outro lado, as intuições são mais estáveis do que os primeiros estudos sugerem, ainda assim a X-Phi teria prestado um serviço inestimável ao forçar a filosofia a examinar criticamente seus próprios fundamentos metodológicos.

Subjacente ao debate sobre as intuições está uma questão metafísica mais profunda sobre o que são, afinal, os conceitos que a filosofia investiga. Eles têm uma estrutura estável que pode ser descoberta pela investigação filosófica, tal como nos diz o realismo conceitual, ou são, em grande parte, construídos pelas teorias e pelos métodos que usamos para estudá-los, tal como nos diz o construtivismo conceitual?

Esta questão atravessa toda a filosofia da linguagem e da mente contemporâneas, e sua relevância para a X-Phi é direta. Se os conceitos são entidades

em que a estrutura é fixada independentemente de qualquer prática teórica, como supõe uma versão forte do realismo conceitual, então é em princípio possível investigá-los empiricamente, verificando se os juízos dos falantes refletem adequadamente essa estrutura. Se, contudo, os conceitos são em parte constituídos pelas práticas de atribuição e revisão que os sustentam, como sugere o construtivismo conceitual de inspiração wittgensteiniana (Wittgenstein, 1953, §§185-242), então os dados sobre variação intuitiva têm um significado completamente diferente; eles não revelam um erro de acesso a um conceito previamente fixado, mas a própria pluralidade das práticas que o constituem.

Esta questão é particularmente aguda para a X-Phi porque seus métodos dependem crucialmente da ideia de que os conceitos do "povo" ou do "senso comum" (*folk concepts*) têm uma estrutura que pode ser empiricamente investigada. Quando Knobe (2008) pergunta se as pessoas consideram um efeito colateral como intencional, ele pressupõe que há um conceito de "intencional" operando na mente dos participantes, e que as respostas ao survey nos dão acesso a esse conceito.

Mas como sabemos que as respostas refletem a competência conceitual dos participantes e não meramente um ruído performático, ou as influências superficiais do contexto, da formulação da pergunta, do cansaço, do desejo de parecer consistente? Esta distinção, emprestada da linguística gerativa (competência linguística e performance), é crucial para interpretar os dados experimentais.

Edouard Machery (2017), em seus trabalhos metodológicos, tem insistido precisamente neste ponto. Para que um dado experimental conte como evidência sobre um conceito, precisamos de uma teoria que nos diga quando um juízo é constitutivo do conceito e quando é um desvio causado pelo contexto. Sem essa teoria, estamos simplesmente colecionando opiniões, não investigando conceitos. Chomsky (1965) introduziu uma distinção para separar o conhecimento gramatical tácito do falante (competência) de suas produções linguísticas concretas, que podem ser distorcidas por fatores externos (performance). Transposta para a X-Phi, a distinção levanta a seguinte questão. Quando um participante emite um juízo em um *survey*, esse juízo manifesta sua competência conceitual ou é um artefato de performance? A resposta depende de condições experimentais que a X-Phi precisa explicitar e controlar, mas também de

uma teoria prévia sobre o que conta como competência conceitual, e essa teoria é, ela mesma, filosófica.

O problema é que não há consenso sobre qual teoria dos conceitos está correta. Diferentes teorias (a teoria clássica, que vê conceitos como condições necessárias e suficientes; a teoria dos protótipos, que os vê como estruturas probabilísticas; a teoria da teoria (ou teoria-teoria), que os vê como mini-teorias; a teoria atomística, que os vê como símbolos com conexões causais) produzem diferentes expectativas sobre o que conta como evidência e como distinguir competência de performance.

A teoria dos protótipos, por exemplo, associada aos trabalhos de Eleanor Rosch (1991), prediz que os juízos de categorização serão graduados e sensíveis a efeitos de tipicidade, o que não seria um ruído a ser filtrado, mas um dado estrutural sobre a organização dos conceitos (Rosch, 1975, p. 192-200). A teoria atomística, por sua vez, prediz que os juízos sobre conceitos são relativamente robustos a variações de contexto, pois o conteúdo dos conceitos seria fixado por relações causais externas, não por representações internas variáveis.

Isto coloca a X-Phi em uma posição delicada. Para justificar seus métodos, ela precisa de uma teoria dos conceitos; mas a escolha da teoria dos conceitos é, ela mesma, uma questão filosófica controversa. Não há como escapar deste círculo, pois a X-Phi precisa fazer filosofia para fazer ciência, e precisa fazer ciência para informar a filosofia. Esta circularidade não constitui um defeito fatal; ela é, antes, expressão da estrutura dialética que caracteriza toda investigação em que teoria e evidência se codeterminam. O que a X-Phi torna explícito é que a relação entre intuições e teorias filosóficas nunca foi linear nem unidirecional; as teorias selecionam quais intuições contam como evidência, e as intuições exercem pressão revisora sobre as teorias. Tornar essa dialética empiricamente informada é, precisamente, a contribuição metodológica central da filosofia experimental.

Essa estrutura é análoga ao que Imre Lakatos (1978) descreveu como o funcionamento dos "programas de pesquisa científica". O núcleo central de uma teoria é protegido por um cinturão de hipóteses auxiliares que absorve as anomalias, enquanto o progresso científico consiste na capacidade de o programa gerar predições novas (Lakatos, 1978, p. 47-52). Na X-Phi, o núcleo correspon-

de às hipóteses filosóficas sobre os conceitos; o cinturão, às teorias cognitivas sobre a formação de juízos.

3. Epistemologia da Filosofia Experimental

Passemos agora às questões epistemológicas. Que tipo de conhecimento os experimentos filosóficos produzem? Como os dados empíricos podem informar a teorização filosófica? Quais são os limites metodológicos da abordagem experimental? A resposta a esta pergunta depende, em parte, de como entendemos a relação entre os juízos das pessoas comuns e os conceitos filosóficos que nos interessam. Uma posição otimista, adotada por muitos filósofos experimentais, é que os juízos do "povo" ou do senso comum nos dão acesso direto à estrutura dos conceitos que a filosofia tradicional sempre investigou. Quando perguntamos a alguém se uma pessoa em uma situação hipotética "sabe" algo, ou se uma ação foi "intencional", ou se alguém é "moralmente responsável", estamos, segundo esta visão, acessando os mesmos conceitos que os filósofos analisam.

Uma posição mais cética, defendida por alguns críticos, é que os juízos das pessoas comuns são apenas isso, juízos comuns, influenciados por todo tipo de fator idiossincrático e irrelevante para a investigação filosófica. O que nos interessaria, segundo estes críticos, não é o que as pessoas pensam sobre conhecimento ou responsabilidade, mas o que o conhecimento e a responsabilidade realmente são. E esta é uma questão normativa, não descritiva. Esta objeção ecoa a distinção, cara à tradição kantiana, entre *quid facti* e *quid juris*. O fato de que os sujeitos formam certos juízos (*quid facti*) não implica que esses juízos sejam normativamente corretos ou que revelem a natureza do que está sendo julgado (*quid juris*). A X-Phi, nesta leitura crítica, responderia adequadamente à questão de fato, mas seria, por princípio, incapaz de responder à questão de direito.

Esta tensão entre o descritivo e o normativo perpassa toda a epistemologia da X-Phi. Os dados experimentais são, inegavelmente, descritivos; eles nos dizem como as pessoas de fato pensam. Mas o objetivo da filosofia é, em grande parte, normativo, i.e., ela quer nos dizer como devemos pensar, quais crenças são justificadas, quais ações são corretas, quais conceitos são adequados.

Como passar do descritivo ao normativo? Esta é a questão central que a X-Phi precisa enfrentar. Uma possibilidade é que os dados descritivos tenham implicações normativas indiretas. Se descobrimos, por exemplo, que as pessoas são sistematicamente influenciadas por fatores irrelevantes em seus juízos morais, isso pode nos levar a desconfiar desses juízos e a buscar procedimentos mais confiáveis. Se descobrimos que as intuições variam transculturalmente, isso pode nos levar a questionar teorias filosóficas que pretendem ser universalmente válidas, mas que, na verdade, apenas refletem intuições locais. Esta passagem indireta do descritivo ao normativo tem estrutura semelhante ao argumento do "equilíbrio reflexivo amplo" de John Rawls (1971). A teoria filosófica não apenas codifica as intuições iniciais dos sujeitos, mas as submete à revisão reflexiva em face de princípios e de evidências sobre suas condições de produção. Se tais condições revelam que certas intuições são produtos de viés, sua autoridade normativa diminui proporcionalmente (Rawls, 1971, p. 48-51).

Outra possibilidade é que os dados descritivos sirvam para refinar nossas teorias normativas, não para substituí-las. Uma teoria do conhecimento que ignore completamente como as pessoas realmente pensam corre o risco de se tornar irrelevante. Uma teoria que leve em conta as intuições comuns pode ser mais bem-sucedida em explicar por que certos casos nos parecem problemáticos.

A X-Phi, por operar na fronteira entre filosofia e psicologia, enfrenta desafios metodológicos peculiares. Ela precisa satisfazer os padrões de rigor de duas disciplinas diferentes, que nem sempre coincidem. Da psicologia, a X-Phi herda a exigência de rigor experimental, tendo de apresentar amostras adequadas, controles apropriados, análise estatística correta, replicação dos resultados, transparência nos métodos. Os estudos que ignoram esses requisitos são justamente criticados como "amadores" do ponto de vista psicológico.

A dificuldade é que os critérios de rigor das duas disciplinas podem entrar em conflito. Capturar com precisão filosófica um conceito técnico exige formulações que podem ser incompreensíveis para participantes leigos. Simplificar a formulação para torná-la acessível arrisca distorcer o conceito de interesse. Esta tensão entre validade externa (generalização) e validade interna (fidelidade conceitual) é constitutiva da X-Phi, e sua gestão é uma das principais habilidades que o filósofo experimental precisa desenvolver.

Da filosofia, a X-Phi herda a exigência de rigor conceitual, i.e., clareza na formulação das perguntas, atenção às distinções sutis, consciência das tradições argumentativas, relevância para os debates estabelecidos. Estudos que ignoram esses requisitos são criticados como irrelevantes ou como simplesmente mudando de assunto.

O desafio é satisfazer ambas as exigências simultaneamente. Uma pergunta de *survey* precisa ser compreensível para participantes leigos e ao mesmo tempo capturar as nuances do conceito filosófico em questão. Uma análise estatística precisa ser correta e ao mesmo tempo informar o debate filosófico de maneira relevante.

David Pizarro (2009), psicólogo e colaborador frequente de filósofos experimentais, sustenta que a X-Phi é, na prática, psicologia social orientada por um repertório mais rico e filosoficamente informado de questões investigativas. O que a psicologia ganha com a colaboração é um horizonte de perguntas que a psicologia tradicional raramente formula; o que a filosofia ganha é o acesso a métodos empíricos capazes de testar hipóteses sobre o funcionamento da mente que a introspecção, por si só, não pode validar ou refutar.

Psicólogos podem avaliar que os filósofos subestimam a complexidade dos desenhos experimentais; filósofos podem avaliar que os psicólogos operam com reconstruções conceptualmente inadequadas dos problemas filosóficos em jogo. O sucesso da colaboração depende de uma disposição de ambas as partes para aprender com a outra e para tolerar o desconforto de operar em uma zona de fronteira onde nenhuma das duas disciplinas tem hegemonia.

Uma vez obtidos os dados, surge a questão talvez mais difícil. Como interpretá-los? O que significa, por exemplo, a descoberta de que as pessoas respondem diferentemente a casos abstratos e concretos sobre livre-arbítrio? Os experimentos de Nichols e Knobe (2008) sobre este tópico são um excelente exemplo. Quando perguntados abstratamente se é possível ser moralmente responsável em um universo determinístico, a maioria dos participantes diz que não, uma resposta incompatibilista. Mas quando apresentados a um caso concreto e emocionalmente carregado, como um homem que mata sua família, a maioria diz que o agente é moralmente responsável, uma resposta compatibilista (Nichols; Knobe, 2007, p. 663-685).

Como interpretar este resultado? Várias interpretações são possíveis. A interpretação psicológica, ou teoria do processo dual, sugere que a mente humana opera com dois sistemas de processamento. Um sistema lento, deliberativo, abstrato, que tende ao incompatibilismo, e um sistema rápido, automático, emocional, que tende ao compatibilismo. Esta leitura é compatível com a teoria dual dos sistemas mentais de Daniel Kahneman (2011), que distingue entre o Sistema 1, rápido, automático, heurístico, e o Sistema 2, lento, deliberativo, normativo (Kahneman, 2011, p. 20-30). Aplicada ao livre-arbítrio, a hipótese seria que o Sistema 1 produz respostas compatibilistas ao detectar intencionalidade e merecimento de punição, enquanto o Sistema 2 produz respostas incompatibilistas ao raciocinar abstratamente sobre a causalidade determinística. O dado revela, assim, a arquitetura da cognição humana.

A interpretação sugere que o dado revela que o conflito filosófico entre compatibilismo e incompatibilismo tem raízes profundas na psicologia humana. As pessoas são, por assim dizer, "naturalmente" divididas sobre esta questão. A filosofia deve levar este fato em conta, seja para arbitrar o conflito, seja para explicar por que ele persiste.

A interpretação cética, ou da ilusão, sugere que o dado mostra que nossos juízos sobre livre-arbítrio são influenciados por fatores emocionais irrelevantes, o que indica que não devemos confiar neles. Talvez a questão do livre-arbítrio seja, em última análise, mal colocada. Essa posição aproxima-se do eliminativismo conceitual de Machery (2017). Se os juízos sobre livre-arbítrio são sistematicamente distorcidos por variáveis emotivas, talvez o conceito de "livre-arbítrio" não seja um conceito coerente, mas antes um amálgama instável de intuições incompatíveis que a filosofia tem tratado, equivocadamente, como se fossem manifestações de um único conceito subjacente (Machery, 2017, p. 131-145).

Cada uma destas interpretações tem implicações diferentes para a filosofia. A primeira informa a psicologia moral, mas não resolve a questão filosófica. A segunda legitima o conflito filosófico como expressão de uma tensão real na cognição humana. A terceira sugere que talvez devêssemos abandonar a questão. O ponto importante é que os dados, por si sós, não escolhem entre estas interpretações. Eles precisam ser interpretados à luz de pressupostos teóricos sobre a mente, sobre a filosofia, sobre a relação entre intuições e teorias. A

epistemologia da X-Phi é, portanto, uma epistemologia da interpretação, e tal tipo de interpretação é sempre uma atividade filosófica.

Um desafio epistemológico adicional para a X-Phi diz respeito à generalização de seus resultados. A maioria dos estudos iniciais foi realizada com participantes ocidentais, educados, industrializados, ricos e democráticos, o que os psicólogos chamam de amostras WEIRD (*Western*, *Educated*, *Industrialized*, *Rich*, *Democratic*).

O problema, como Henrich, Heine & Norenzayan (2010) mostraram de forma contundente, é que as amostras WEIRD são atípicas em termos psicológicos. Em uma série de estudos, eles demonstraram que participantes ocidentais diferem sistematicamente de participantes não ocidentais em uma ampla gama de tarefas cognitivas, incluindo aquelas relevantes para a filosofia, como percepção visual, raciocínio moral, noções de justiça. Isto coloca um problema sério para a X-Phi. Se seus resultados são baseados principalmente em amostras WEIRD, em que medida eles podem ser generalizados para a espécie humana como um todo? Uma filosofia que pretende falar sobre conceitos humanos universais não pode se dar ao luxo de ignorar a diversidade humana (Henrich; Heine; Norenzayan, 2010, p. 61-83).

A resposta da X-Phi a este desafio tem sido dupla. Por um lado, tem havido um esforço crescente para realizar estudos transculturais, incluindo as pesquisas pioneiras de Machery (2017) e colaboradores sobre referência de nomes próprios. Por outro lado, alguns filósofos experimentais argumentam que a diversidade cultural é, ela mesma, um dado filosófico relevante. Se diferentes culturas convergem em certas intuições fundamentais, isso é evidência a favor de que essas intuições rastreiam algo real sobre os conceitos em questão, ou sobre a arquitetura cognitiva humana compartilhada. Se divergem, isso pode ser evidência de que os conceitos são parcialmente constituídos por práticas culturais particulares, o que tem implicações diretas para qualquer pretensão de universalidade das teorias filosóficas.

O desafio da amostragem, longe de ser um problema técnico menor, toca no coração da epistemologia da X-Phi. Ele nos lembra que a filosofia, mesmo em sua vertente experimental, precisa levar a sério a diversidade humana e evitar a tentação de universalizar o particular.

4. Ética da Filosofia Experimental

Chegamos às questões éticas. Que responsabilidades os filósofos experimentais têm para com os participantes de suas pesquisas? Como lidar com as implicações sociais de seus achados? Que cuidado devem ter ao comunicar resultados ao público? A X-Phi, ao utilizar métodos empíricos, herda as obrigações éticas da pesquisa em psicologia e ciências sociais. Isto significa que seus praticantes devem seguir protocolos estabelecidos para a pesquisa com seres humanos, como obter consentimento informado, proteger a privacidade dos participantes, evitar causar danos, e ser transparentes sobre os procedimentos. Essas exigências podem parecer óbvias, mas têm implicações práticas importantes. A formulação das perguntas em um survey não é apenas uma questão metodológica, mas também ética. Perguntas mal formuladas podem induzir respostas, criar confusão, ou mesmo causar desconforto psicológico desnecessário. O filósofo experimental tem a responsabilidade de minimizar estes riscos.

Além disso, a X-Phi frequentemente lida com tópicos sensíveis, como moralidade, religião, política e sexualidade. Os participantes podem ser convidados a fazer juízos sobre casos hipotéticos que envolvem violência, traição, discriminação ou outros temas perturbadores. É responsabilidade do pesquisador garantir que os participantes estejam cientes da natureza do estudo e que possam desistir a qualquer momento sem prejuízo. Estas considerações éticas, que são rotineiras na psicologia, podem ser novidade para filósofos acostumados apenas à pesquisa de poltrona. A X-Phi, neste sentido, exige não apenas novas habilidades metodológicas, mas também uma nova consciência ética.

A questão da representatividade das amostras, discutida anteriormente em termos epistemológicos, tem também uma dimensão ética crucial. Quando a X-Phi baseia suas conclusões em amostras WEIRD, ela corre o risco de cometer uma injustiça epistêmica no sentido de Miranda Fricker (2007), pois desconsidera ou marginaliza as perspectivas de grupos não ocidentais, não educados, não ricos (Fricker, 2007, p. 147-161). Uma X-Phi que generalize a partir de amostras WEIRD pode perpetuar precisamente essa assimetria, ao tratar os conceitos e as intuições de grupos específicos como se fossem universalmente humanos. Essa injustiça é particularmente problemática porque a X-Phi frequentemente faz afirmações sobre a cognição humana em geral. Se estas afirmações são

baseadas apenas em um subconjunto restrito da humanidade, elas são, no mínimo, incompletas. Mais seriamente, elas podem servir para legitimar uma visão de mundo ocidental como se fosse universal, contribuindo para a perpetuação de hierarquias epistêmicas.

Os estudos de Jesse Prinz (2010) sobre diferenças de gênero em juízos morais ilustram tanto o potencial quanto os riscos deste tipo de investigação. Por um lado, eles chamam a atenção para diferenças reais que podem ter implicações importantes, como na composição de tribunais. Por outro lado, eles podem ser mal interpretados como justificando estereótipos de gênero ou excluindo mulheres de certos papéis.

A discussão sobre a composição de gênero em tribunais superiores insere-se em uma tradição mais ampla que investiga a relação entre perspectiva epistêmica e julgamento institucional. A hipótese de que diferentes posições sociais engendram distintas sensibilidades morais encontra fundamento na epistemologia feminista, especialmente na noção de *standpoint epistemology*. Essa concepção, desenvolvida por Sandra Harding (1991), sustenta que a posição social do sujeito cognoscente possui relevância epistêmica e não constitui um elemento a ser neutralizado em nome de uma objetividade abstrata (Harding, 1991, p. 119-137).

A responsabilidade do filósofo experimental, neste contexto, é dupla. Ele deve buscar ativamente amostras diversas e evitar generalizações apressadas baseadas em dados limitados. E deve comunicar seus resultados com cuidado, explicando o que eles mostram e, crucialmente, o que eles não mostram. A X-Phi não pode controlar como seus resultados serão usados por outros, mas tem a responsabilidade de apresentá-los de forma que minimize o risco de má interpretação. O caso de Prinz (2010) acerca da composição de gênero em tribunais é ilustrativo da tensão entre o potencial emancipatório e o risco de instrumentalização dos resultados da X-Phi. Se diferenças sistemáticas de gênero se manifestam em decisões judiciais, então a composição do tribunal tem consequências substantivas para a justiça, o que pode informar tanto políticas de nomeação quanto a crítica de práticas institucionais arraigadas.

Outro exemplo são os estudos sobre a influência de fatores emocionais em juízos morais. Se as pessoas são mais propensas a punir quando estão com raiva, isso pode nos levar a questionar a confiabilidade de julgamentos em con-

textos judiciais ou a buscar procedimentos que minimizem a influência de emoções irrelevantes.

Essas implicações práticas colocam a X-Phi diante de uma questão ética fundamental. Que responsabilidade os filósofos experimentais têm pelas consequências sociais de seus achados? Não há resposta fácil para esta pergunta. Por um lado, os pesquisadores não podem controlar como seus resultados serão usados. Por outro lado, ignorar completamente as possíveis consequências seria uma forma de irresponsabilidade. A posição mais defensável parece ser aquela que Stephen Turner (2003) denomina "responsabilidade epistêmica". O pesquisador que produz conhecimento com potencial de impacto social tem a obrigação de acompanhar esse conhecimento com os recursos interpretativos necessários para que sua recepção pública seja adequada, sem que isso implique a supressão de resultados ou a autocensura dos procedimentos de investigação.

A circulação midiática de resultados da X-Phi tende a operar com reduções severas, pois suas afirmações nuançadas sobre variações intuitivas em determinados contextos experimentais são transformadas em declarações gerais sobre a irracionalidade humana ou a relatividade da moral. Esta atenção traz consigo uma responsabilidade especial. O filósofo experimental, mais do que o filósofo tradicional, precisa ser um comunicador público competente. Ele precisa explicar não apenas o que seus resultados mostram, mas também suas limitações, o contexto em que foram obtidos, e as cautelas necessárias para interpretá-los. Isto é particularmente importante quando os resultados tocam em questões socialmente sensíveis. Estudos sobre diferenças de gênero, raça ou cultura em juízos morais ou cognitivos podem facilmente ser mal interpretados e usados para fins questionáveis. O filósofo experimental tem a responsabilidade de antecipar estas possibilidades e de comunicar seus resultados de forma a minimizar os possíveis danos.

Além disso, a X-Phi pode desempenhar um papel educativo importante. Ao mostrar como a cognição humana funciona, com todos os seus vieses e limitações, ela pode ajudar o público a se tornar mais consciente de suas próprias tendências e a buscar procedimentos mais confiáveis para a formação de juízos. Esta função educativa é, ela mesma, uma contribuição ética significativa.

5. Considerações Finais

A filosofia experimental não é apenas mais uma disciplina no espectro filosófico. Ela é, de forma peculiar, uma autorreflexão da filosofia sobre seus próprios métodos e pressupostos. Ao aplicar métodos empíricos a questões tradicionalmente filosóficas, ela nos força a perguntar de onde vêm nossas intuições, quão confiáveis elas são, como elas variam entre culturas, grupos sociais, indivíduos, e, finalmente, o que isso significa para as teorias que construímos sobre elas.

Metafilosoficamente, vimos que a X-Phi desafia a imagem tradicional da filosofia como disciplina puramente *a priori*, baseada apenas em razão e argumento. Ela propõe que as intuições, que sempre desempenharam um papel central no método filosófico, podem e devem ser investigadas empiricamente. Isto não significa abandonar a análise conceitual, mas sim informá-la com dados sobre como as pessoas realmente pensam.

Epistemologicamente, vimos que a interpretação dos dados experimentais é complexa e envolve pressupostos teóricos profundos. Não basta coletar respostas; é preciso entender o que elas significam, se refletem competência conceitual ou ruído performático e se são generalizáveis para além da amostra estudada. A diversidade cultural e a variação das intuições colocam desafios sérios para qualquer pretensão de universalidade.

Eticamente, vimos que a X-Phi carrega responsabilidades especiais. Seus resultados podem ter implicações sociais profundas e podem ser mal interpretados. Os pesquisadores têm o dever de proteger os participantes, de buscar amostras diversas, de comunicar seus resultados com cuidado e nuance.

O fio que atravessa esses três domínios, metafilosófico, epistemológico e ético, é a questão da reflexividade. A X-Phi é, em última instância, a filosofia tornando-se objeto de sua própria investigação. Aplicando a si mesma os padrões de escrutínio crítico que ela exige das ciências que fundamenta. Esta reflexividade não é um gesto cético nem autodestrutivo; é a condição de possibilidade de uma prática filosófica madura, capaz de reconhecer suas próprias contingências históricas, cognitivas e sociais sem abdicar de suas pretensões normativas.

O pluralismo que emerge destas reflexões reconhece que não há um método filosófico único e privilegiado. A análise conceitual tradicional, a experimentação controlada, os estudos transculturais, a neuroimagem, cada um desses métodos ilumina diferentes aspectos dos problemas filosóficos. Nenhum tem a palavra final. A tarefa do filósofo reflexivo é saber quando e como usar cada ferramenta, e estar consciente de suas limitações. Embora essa posição seja metodologicamente pluralista, ela não é indiferente. O pluralismo não implica que todos os métodos são igualmente adequados para todas as questões, mas que a adequação metodológica é ela mesma uma questão que precisa ser respondida caso a caso, à luz dos objetivos da investigação e das condições em que ela é conduzida. O que a X-Phi exige, neste sentido, é uma metacognição metodológica, uma consciência explícita das condições de validade e dos limites de cada instrumento investigativo.

A filosofia experimental nos lembra de algo que este livro inteiro tentou mostrar, a saber, que a filosofia não é uma atividade desencarnada, pairando acima das ciências que examina. A filosofia é praticada por cérebros humanos, em contextos históricos e sociais específicos, sujeita às mesmas limitações cognitivas que as ciências que ela estuda. Compreender essas limitações não é diminuir a filosofia (ou qualquer outra ciência); é torná-la mais consciente de si mesma, mais rigorosa, mais humilde.

As conexões com os capítulos anteriores são múltiplas. A discussão sobre intuições na X-Phi ecoa as discussões sobre o papel das intuições na linguística e na ética. A preocupação com a diversidade cultural ecoa os debates na antropologia sobre o relativismo e o universalismo. A atenção aos métodos empíricos conecta-se com a ênfase na experimentação na psicologia e na ciência cognitiva. A reflexão sobre a confiabilidade dos juízos ecoa as discussões sobre testemunho e autoridade na epistemologia social e na ciência da informação.

A filosofia experimental é, em certo sentido, a filosofia se tornando consciente de sua própria contingência histórica e cognitiva. É a filosofia se perguntando o que significa filosofar, para criaturas como nós, com cérebros como os nossos, em culturas como as nossas. Esta pergunta tem uma dimensão que poderíamos chamar de transcendental, no sentido kantiano. Em outros termos, ela não questiona apenas os conteúdos do saber filosófico, mas as condições de possibilidade da própria atividade filosófica, as estruturas cognitivas, culturais

e históricas que tornam possível, e ao mesmo tempo limitam, o pensamento filosófico humano. Neste sentido, a X-Phi não é uma ameaça à filosofia, mas sua mais radical exigência de autossuperação. E essa pergunta, longe de ser uma ameaça, é talvez a mais filosófica de todas.

Considerações Finais

1. O Percurso

Chegamos ao fim. Mas "fim" é uma palavra enganosa. Uma jornada filosófica não termina quando viramos a última página; ela continua na cabeça do leitor, nas perguntas que ficaram sem resposta, nas conexões que ele mesmo fará com suas próprias experiências e preocupações. Se este livro funcionou como esperado, o leitor não está agora com um conjunto de respostas definitivas, mas com um repertório mais rico de perguntas, uma sensibilidade mais aguçada para as dimensões filosóficas das ciências que estuda ou pratica.

Atravessamos um território vasto, que se estende da antropologia à filosofia experimental. No caminho, visitamos a linguística, a matemática, a ciência da informação, a arquivologia, a biblioteconomia, a museologia, o turismo, a engenharia, a ciência da computação, a criptografia, a ciência cognitiva, a robótica, a inteligência artificial e a arte digital. Em cada parada, um mesmo fenômeno se repetia, pois as ciências que examinamos não operam no vácuo; elas fazem afirmações sobre o mundo que são informadas por descobertas empíricas, mas também incorporam pressupostos filosóficos profundos que raramente são explicitados.

A antropologia nos colocou diante da cultura como fenômeno real, múltiplo, emergente e histórico. O dilema entre universalismo e particularismo se impôs como constitutivo da disciplina, pois ela precisa de algum universalismo para ser ciência comparativa, mas também de algum particularismo para cumprir sua função crítica de desnaturalizar nossas próprias categorias. A linguística, por sua vez, apresentou a língua como fenômeno simultaneamente cognitivo, social e formal. O debate entre platonismo linguístico, nominalismo behaviorista, psicologismo mentalista e institucionalismo social deixou claro que não há uma resposta única para o que é uma língua, mas sim diferentes aspectos que diferentes teorias iluminam. A matemática nos introduziu no estranho reino dos números, habitantes de uma zona fronteiriça entre o abstrato e o concreto. Novamente, o dilema entre platonismo e nominalismo, que atravessa milênios de filosofia,

não encontra solução definitiva, embora o naturalismo matemático modesto ofereça um caminho para viver produtivamente com essa tensão.

A ciência da informação, por seu turno, expôs a natureza dual da informação, física e semântica, e o realismo informacional modesto reconheceu que a informação é real, mas seu modo de existência é complexo e irredutível a uma única categoria. A arquivologia ensinou que documentos são entidades que persistem através de mudanças materiais graças a cadeias de custódia e práticas institucionais. A autenticidade se revelou uma propriedade socialmente sustentada, dependente de infraestruturas epistêmicas e jurídicas. A biblioteconomia mostrou que livros são simultaneamente tipos abstratos e tokens materiais, e que bibliotecas funcionam como máquinas de pensar em escala coletiva. A museologia evidenciou que objetos se transformam ao entrar em coleções, adquirindo novas camadas de significado, tornando-se semióforos que conectam o visível ao invisível. No turismo, descobrimos que lugares se tornam destinos quando enquadrados pelo olhar do visitante. A autenticidade, tão central para o discurso turístico, mostrou sua natureza plural, podendo ser objetiva, construtiva ou existencial dependendo do contexto.

A engenharia nos confrontou com artefatos técnicos que carregam valores, cristalizam relações de poder, materializam visões de mundo. A ciência da computação expôs a opacidade dos algoritmos, os limites da verificabilidade formal, as barreiras que a complexidade computacional impõe ao conhecimento. A criptografia desvendou a natureza delicada do segredo, mostrando como a confiança pode ser distribuída sem autoridades centrais através de matemática engenhosa, e evidenciou que a privacidade não é um luxo mas um campo de batalha político. A ciência cognitiva nos colocou diante do problema difícil da consciência, lembrando que a experiência subjetiva pode resistir a explicações puramente funcionais ou físicas. A robótica e a inteligência artificial levantaram questões éticas urgentes sobre responsabilidade, autonomia e dignidade na criação de mentes artificiais. E a arte digital, que se utiliza dos artefatos técnicos dos engenheiros, de registros descentralizados de informação e de inteligências artificiais, revolucionou as noções de autoria, autenticidade, aura e valor, mostrando como a criação artística se transforma no ambiente digital.

A filosofia experimental, por fim, submeteu a própria filosofia ao escrutínio empírico, investigando as intuições que fundamentam nossas teorias.

2. O Fio Condutor

Se há um fio que une todos esses capítulos, ele talvez seja a tríade informação-memória-técnica, à qual podemos acrescentar um quarto elemento, a criação. A antropologia investiga a cultura como meio pelo qual os humanos criam e transmitem informação. A linguística estuda a linguagem como veículo dessa transmissão. A matemática fornece a estrutura abstrata que torna possível a própria ideia de informação.

Por sua vez, a ciência da informação examina o que acontece quando a linguagem se torna registro durável. A arquivologia se debruça sobre como esses registros são preservados como prova. A biblioteconomia analisa como são organizados para o acesso. A museologia investiga como os objetos se tornam portadores de memória cultural. O turismo explora como essa memória é experienciada.

A engenharia, então estuda como construímos artefatos que transformam o mundo. A ciência da computação examina como processamos a informação em escala e velocidade inéditas. A criptografia investiga como protegemos essa informação. A ciência cognitiva analisa como a mente processa a informação. A robótica e a IA exploram como podemos criar sistemas que processam informação e agem no mundo. A arte digital estuda como podemos criar novos objetos estéticos nesse ambiente informacional.

E, finalmente, a filosofia experimental investiga os processos cognitivos que estão na base de toda essa atividade.

Informação, memória, técnica e criação são quatro conceitos que resumem a aventura humana. Criamos informação para dar sentido ao mundo. Preservamos a memória para que o passado não se perca. Desenvolvemos técnicas para transformar o presente e construir o futuro. Criamos arte para expressar e experienciar o que escapa à mera funcionalidade. As ciências que examinamos são maneiras diferentes de participar dessa aventura. Compreender filosoficamente essas ciências é, de fato, compreender a nós mesmos.

3. O Pluralismo como Estratégia

Ao longo do livro, uma palavra apareceu repetidamente, a saber, “pluralismo”. Pluralismo ontológico na antropologia, pluralismo metodológico na linguís-

tica, pluralismo conceitual na ciência da informação, pluralismo ético em quase todas as disciplinas. Convém, neste momento final, esclarecer o que esse pluralismo significa e, igualmente importante, o que ele não significa. E, mais honestamente ainda, é preciso reconhecer o seu caráter estratégico.

O pluralismo que adotei não é uma posição metafísica profunda sobre a estrutura última da realidade. Não estou afirmando que o mundo é, em sua essência, plural, ou que diferentes teorias incompatíveis podem ser simultaneamente verdadeiras num sentido forte do termo. Essas são questões filosóficas profundas, que envolvem debates sobre realismo, relativismo, verdade e objetividade, debates que têm séculos e que não pretendi resolver aqui.

O pluralismo deste livro é, antes de tudo, uma estratégia *pedagógica* e uma postura *pragmática* diante da diversidade das ciências que examinamos. Cada uma dessas áreas tem suas próprias tradições, seus próprios métodos, seus próprios problemas. Profissionais formados em diferentes escolas teóricas podem e devem dialogar, podem e devem aprender uns com os outros, sem que isso signifique que todos precisam concordar sobre os fundamentos últimos de suas disciplinas. A ideia deste livro é mais estimular o debate do que determinar os fundamentos de cada disciplina estudada.

Um arquivista formado na tradição diplomática pode ter compromissos ontológicos diferentes de um arquivista que trabalha com documentos digitais e cadeias de custódia baseadas em blockchain. Um linguista gerativista e um linguista funcionalista partem de pressupostos diferentes sobre a natureza da linguagem. Um museólogo que vê os objetos como semióforos, nos termos de Pomian (1984), tem uma visão diferente daquele que enfatiza a materialidade e a agência dos objetos. Um engenheiro que projeta pontes enfrenta questões diferentes de um engenheiro de software que desenvolve algoritmos de IA.

O pluralismo que percorre o livro é uma tentativa de criar um espaço onde essas diferentes perspectivas possam coexistir produtivamente, onde os profissionais possam fazer escolhas informadas filosoficamente sem serem forçados a adotar uma posição única e exclusiva. Não se trata de dizer que todas as teorias são igualmente válidas, que qualquer posição é tão boa quanto qualquer outra. Trata-se, sim, de reconhecer que diferentes problemas podem exigir diferentes ferramentas conceituais, que diferentes contextos podem demandar

diferentes ênfases, e que a escolha entre teorias rivais envolve julgamentos complexos que não se reduzem a algoritmos simples.

É claro que essa posição pluralista pode não se sustentar para além de sua função pragmática. É possível que, em um nível mais profundo de investigação, algumas teorias se mostrem mais adequadas que outras, que alguns pressupostos ontológicos sejam mais defensáveis que outros, que algumas respostas aos problemas filosóficos das diferentes áreas sejam corretas e outras erradas. A discussão filosófica é profunda e exige rigor. Mas este livro nunca pretendeu oferecer respostas definitivas. Sua pretensão era mais modesta, a saber, a de levantar questões importantes para as diversas áreas, pensar com clareza sobre elas, com sistematicidade, com lógica, mas também com sensibilidade para a complexidade e a diversidade das práticas científicas.

Se o pluralismo funcionou como uma ferramenta para abrir espaço para essa conversa, já cumpriu seu papel. Que cada leitor, a partir de sua própria área, de suas próprias convicções, de suas próprias perguntas, possa levar adiante a investigação e aprofundar as questões que ficaram em aberto.

4. Tensões Recorrentes

Em cada ciência que visitamos, encontramos tensões que não se resolvem facilmente. Tensões entre o universal e o particular, entre o objetivo e o subjetivo, entre o descritivo e o normativo, entre o abstrato e o concreto, entre o puro e o aplicado. Essas tensões não são defeitos a serem superados, mas características constitutivas de ciências que estudam fenômenos complexos.

A antropologia vive a tensão entre universalismo e particularismo. Se ignorasse o particular em nome do universal, perderia sua função crítica de desnaturalizar nossas próprias categorias. A linguística enfrenta a tensão entre rigor formal e sensibilidade contextual. Se desconsiderasse a diversidade em nome dos universais, empobreceria a compreensão da riqueza das línguas humanas. A matemática habita a tensão entre o abstrato e o concreto. Se negligenciasse o abstrato em nome do concreto, perderia o que há de mais fascinante em seus objetos.

A ciência da informação lida com a tensão entre o físico e o semântico. Se ignorasse o significado em nome dos bits, perderia exatamente o que torna a

informação interessante. A arquivologia enfrenta a tensão entre materialidade e conteúdo. Se desprezasse a materialidade em nome do conteúdo, comprometeria a capacidade de autenticar documentos. A biblioteconomia vive a tensão entre o tipo abstrato e o *token* material. Se abandonasse a materialidade em nome do texto, perderia a dimensão histórica e estética da cultura letrada. A museologia se debate com a tensão entre o objeto material e o semióforo. Se desconsiderasse o significado em nome da materialidade, impossibilitaria contar histórias com objetos. O turismo enfrenta a tensão entre experiência autêntica e mediação. Se ignorasse a mediação em nome da experiência autêntica, perderia de vista como as experiências são construídas.

Por sua vez, a engenharia lida com a tensão entre eficiência técnica e valores éticos. Se negligenciasse os valores em nome da eficiência, perderia de vista as dimensões políticas da tecnologia. A ciência da computação vive a tensão entre eficiência algorítmica e transparência. Se desprezasse a opacidade em nome da eficiência, comprometeria a explicabilidade e a prestação de contas. A criptografia enfrenta a tensão entre privacidade individual e segurança coletiva. Se ignorasse a ética em nome da segurança técnica, perderia de vista as implicações sociais da vigilância. A ciência cognitiva se debate com a tensão entre objetividade científica e subjetividade experiencial. Se desconsiderasse a subjetividade em nome da objetividade, perderia o que torna a consciência um problema tão fascinante. A robótica lida com a tensão entre autonomia e controle. Se negligenciasse a corporificação em nome da inteligência abstrata, perderia de vista como o corpo molda a cognição. A arte digital enfrenta a tensão entre autoria humana e criação algorítmica. Se desprezasse a materialidade dos sistemas em nome da experiência estética, perderia de vista como a infraestrutura técnica molda a criação.

E a filosofia experimental vive a tensão entre descrição empírica e normatividade filosófica. Se ignorasse a normatividade em nome da descrição, perderia o que torna a filosofia filosofia.

O que emerge dessa constatação não é um ceticismo paralisante, mas uma compreensão mais matizada da atividade científica. As ciências não são sistemas fechados de verdades eternas; são práticas humanas, históricas, situadas, que lidam com fenômenos complexos demais para serem facilmente capturados por uma única perspectiva. A tarefa do cientista reflexivo, do profissional consci-

ente, é navegar essas tensões com olhos abertos, fazendo escolhas informadas, assumindo responsabilidades, reconhecendo limites.

5. Filosofia como Cartografia

Se tivesse que resumir em uma imagem o que este livro tentou fazer, eu diria que ele é um exercício de cartografia conceitual. Não uma cartografia que pretende mapear todo o território de uma vez, com precisão milimétrica, mas uma cartografia que traça algumas rotas, assinala alguns perigos, indica algumas direções possíveis.

Cada capítulo foi uma expedição a um território diferente, seja o território da cultura, da linguagem, dos números, da informação, dos documentos, dos livros, dos objetos museais, das experiências turísticas, dos artefatos técnicos, dos algoritmos, dos segredos criptográficos, da mente, dos robôs, das obras digitais, das intuições filosóficas. Em cada expedição, procuramos identificar as principais cadeias de montanha (os problemas filosóficos centrais), os principais rios (as teorias que tentam navegar esses problemas), as zonas de conflito (os debates entre posições rivais) e as áreas de sombra (as questões que permanecem em aberto).

Não construímos estradas definitivas. Não declaramos nenhum território completamente explorado. Não oferecemos mapas que pretendem ser a última palavra. O que oferecemos são instrumentos de navegação, ferramentas conceituais que cada leitor pode usar em suas próprias expedições, em seus próprios territórios, com seus próprios problemas.

A filosofia como cartografia não é uma filosofia que dá respostas; é uma filosofia que ajuda a formular melhor as perguntas. Não é uma filosofia que diz o que fazer; é uma filosofia que ajuda a entender o que está em jogo quando escolhemos fazer uma coisa em vez de outra. Não é uma filosofia que elimina a incerteza; é uma filosofia que nos ajuda a navegar a incerteza com os olhos abertos.

6. O Papel do Leitor

Este livro termina, mas as questões que ele levantou permanecem. O que é uma cultura? O que é uma língua? O que é um número? O que é informação? O

que é um documento? O que é um livro? O que é um objeto museal? O que é um destino turístico? O que é um artefato? O que é a computação? O que é o segredo? O que é a mente? O que é a inteligência? O que é a arte na era digital? O que é a filosofia?

Cada leitor levará essas perguntas consigo, e as responderá à sua maneira, informado por suas próprias experiências, suas próprias práticas, seus próprios valores. O arquivista as trará para o seu trabalho com documentos, interrogando-se sobre autenticidade e proveniência. O bibliotecário as aplicará à organização de suas coleções, refletindo sobre classificação e acesso. O museólogo as tomará como guia para a curadoria de suas exposições, pensando na transformação de objetos em semióforos. O profissional do turismo as considerará no planejamento de suas atividades, ponderando sobre autenticidade e hospitalidade. O engenheiro as trará para o design de seus artefatos, questionando-se sobre valores e responsabilidade. O cientista da computação as aplicará ao desenvolvimento de seus algoritmos, refletindo sobre opacidade e explicabilidade. O criptógrafo as considerará na construção de seus sistemas, ponderando sobre privacidade e vigilância. O cientista cognitivo as levará para suas pesquisas sobre a mente, interrogando-se sobre consciência e cognição. O roboticista as trará para a criação de seus robôs, pensando sobre agência e responsabilidade. O artista digital as aplicará à sua prática criativa, refletindo sobre autoria e autenticidade. O filósofo as tomará como matéria para sua reflexão sobre intuições e métodos.

Se este livro contribuiu para que essas respostas sejam um pouco mais refletidas, um pouco mais informadas, um pouco mais abertas à complexidade do mundo, terá cumprido seu propósito. A jornada continua. Em cada arquivo, em cada biblioteca, em cada museu, em cada viagem, em cada projeto de engenharia, em cada linha de código, em cada mensagem criptografada, em cada reflexão sobre a mente, em cada interação com um robô, em cada obra de arte digital, em cada experimento filosófico, as perguntas estarão lá, esperando para serem feitas de novo.

A filosofia não é um ponto de chegada. É um modo de viajar.

7. O Que Permanece

No final de uma longa jornada, o que permanece? Permanece a convicção de que as atividades aparentemente técnicas de organizar, preservar, interpretar, construir e programar são, na verdade, formas de prática filosófica materializada. Elas envolvem, a cada decisão, respostas implícitas ou explícitas às questões mais antigas da filosofia, sobre o que existe, como sabemos e o que devemos fazer.

Permanece a percepção de que as ciências não são ilhas isoladas, mas territórios conectados por problemas comuns. A indução que desafia a filosofia da ciência é a mesma que desafia o aprendizado de máquina. A demarcação que preocupa os filósofos é a mesma que preocupa os profissionais que precisam distinguir ciência de pseudociência em suas áreas. O realismo que divide os filósofos da matemática é o mesmo que divide os cientistas da informação sobre a natureza de seus objetos. A mudança científica que Kuhn (1962) descreveu para a física é a mesma que os arquivistas enfrentam quando seus paradigmas de gestão documental se transformam.

Permanece a compreensão de que as tensões não são defeitos, mas características constitutivas. Todas essas tensões são expressões locais de dilemas mais profundos que atravessam a própria condição humana.

Permanece, por fim, um convite. O convite para que cada leitor, em seu próprio campo, em sua própria prática, leve adiante a reflexão. Para que não se contente com respostas prontas, com manuais, com protocolos estabelecidos. Para que pergunte, sempre que possível, o que está em jogo aqui. Que pressupostos estou assumindo? Que valores estão orientando minha decisão? Que alternativas existem? Que consequências minhas escolhas terão?

A filosofia das ciências, vista assim, não é um exercício de especialistas afastado das preocupações do mundo. É, antes, um engajamento com os fundamentos da nossa vida coletiva na era da informação, da tecnologia e da criação digital. É reconhecer que, quando debatemos padrões de metadados, estamos decidindo que aspectos da realidade serão recuperáveis no futuro. Quando definimos critérios de avaliação documental, estamos determinando quais vozes serão ouvidas pela história. Quando projetamos algoritmos de recomendação, estamos moldando os horizontes intelectuais de milhões de pes-

soas. Quando construímos sistemas de IA, estamos criando entidades que podem um dia reivindicar status moral. Quando criamos arte digital, estamos redefinindo o que significa criar, autenticar, experienciar esteticamente algo.

A jornada deste livro termina aqui, mas as perguntas que ela suscitou continuarão a nos acompanhar. Pois a filosofia, como a ciência, como a tecnologia, como a arte, como a própria vida, é um organismo em crescimento, sempre inacabada, sempre aberta a novas questões, sempre pronta para recomeçar. No fundo, toda essa discussão nunca foi somente sobre as disciplinas. Foi sobre nós. Sobre como nós, antropólogos, linguistas, matemáticos, cientistas da informação, arquivistas, bibliotecários, museólogos, turismólogos, engenheiros, cientistas da computação, criptógrafos, roboticistas, artistas digitais, filósofos, cada um no seu canto, lidamos com perguntas que são maiores do que a nossa área. Se este livro ajudar alguém a perceber que aquela dúvida incômoda que ele tem no dia a dia não é sinal de incompetência, mas sinal de que ele está pensando no lugar certo, já terá valido a pena.

Boa jornada.

Referências

AARONSON, Scott. *Quantum Computing Since Democritus*. Cambridge: Cambridge University Press, 2013.

ABU-LUGHOD, Lila. Writing Against Culture. In: FOX, Richard G. (org.). *Recapturing Anthropology*. Santa Fe: School of American Research Press, 1991. p. 137-162.

ALA — AMERICAN LIBRARY ASSOCIATION. *Code of Ethics*. Chicago: ALA, 1995.

ALA — AMERICAN LIBRARY ASSOCIATION. *Information Literacy: Final Report*. Chicago: ALA, 1989.

ALA — AMERICAN LIBRARY ASSOCIATION. *Library Bill of Rights*. Chicago: ALA, 1939/2019.

AMES, Michael M. *Cannibal Tours and Glass Boxes: The Anthropology of Museums*. Vancouver: UBC Press, 1992.

ANGWIN, Julia et al. *Machine Bias*. ProPublica, 23 maio 2016. Disponível em: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

APPIAH, Kwame Anthony. *Cosmopolitanism: Ethics in a World of Strangers*. New York: W.W. Norton, 2006.

ARENDT, Hannah. *The Human Condition*. Chicago: University of Chicago Press, 1958.

ARISTÓTELES. *Ética a Nicômaco*. Tradução de Leonel Vallandro e Gerd Bornheim. São Paulo: Abril Cultural, 1973. (Os Pensadores).

ARISTÓTELES. *Metafísica*. Tradução de Giovanni Reale. São Paulo: Loyola, 2002.

ARKIN, Ronald C. *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*. Atlanta: Georgia Institute of Technology, 2009.

ASARO, Peter. On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making. *International Review of the Red Cross*, v. 94, n. 886, p. 687-709, 2012.

AUSTIN, John Langshaw. *How to Do Things with Words*. Oxford: Clarendon Press, 1962.

AUTOR, David. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, v. 29, n. 3, p. 3-30, 2015.

AYER, Alfred Jules. Language, Truth and Logic. New York: Dover, 1936/1952.

BACON, Francis. *Novum Organum*. Tradução de José Aluysio Reis de Andrade. São Paulo: Nova Cultural, 1620/2014.

BAKER, Lynne Rudder. *The Metaphysics of Everyday Objects*. Oxford: Oxford University Press, 2007.

BALAGUER, Mark. *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press, 1998.

BANN, Stephen. *The Clothing of Clio: A Study of the Representation of History in Nineteenth-Century Britain and France*. Cambridge: Cambridge University Press, 1984.

BARTHES, Roland. *A Câmara Clara: nota sobre a fotografia*. Rio de Janeiro: Nova Fronteira, 1980/1984.

BATTLES, Matthew. *Library: An Unquiet History*. Nova York: Norton, 2003.

BAUDRILLARD, Jean. *Simulacres et simulation*. Paris: Galilée, 1981.

BAUMAN, Zygmunt. *Globalization: The Human Consequences*. New York: Columbia University Press, 1998.

BEAUCHAMP, Tom; CHILDRESS, James. *Principles of Biomedical Ethics*. 8. ed. Oxford: Oxford University Press, 1979/2019.

BEDAU, Mark A. Weak Emergence. *Noûs*, v. 31, supl. 11, p. 375-399, 1997.

BENACERRAF, Paul. Mathematical Truth. *Journal of Philosophy*, v. 70, n. 19, p. 661-679, 1973.

BENACERRAF, Paul. What Numbers Could Not Be. *Philosophical Review*, v. 74, n. 1, p. 47-73, 1965.

BENEDICT, Ruth. *Patterns of Culture*. Boston: Houghton Mifflin, 1934.

BENHABIB, Seyla. *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton: Princeton University Press, 2002.

BENJAMIN, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press, 2019.

BENJAMIN, Walter. A obra de arte na era de sua reprodutibilidade técnica. In: *Magia e técnica, arte e política*. 7. ed. São Paulo: Brasiliense, 1994. p. 165-196.

BENTHAM, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1789/1996.

BEREITER, Carl. *Education and Mind in the Knowledge Age*. Mahwah: Lawrence Erlbaum Associates, 2002.

BERNSTEIN, Richard J. *Beyond Objectivism and Relativism*. Philadelphia: University of Pennsylvania Press, 1983.

BIANCHI, Raoul. The Political Economy of Tourism Development: A Critical Review. *Anatolia: An International Journal of Tourism and Hospitality Research*, v. 20, n. 2, p. 484-508, 2009.

BIJKER, Wiebe E.; HUGHES, Thomas P.; PINCH, Trevor J. (eds.). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge: MIT Press, 1987.

BLOCK, Ned. Troubles with Functionalism. Minnesota Studies in the Philosophy of *Science*, v. 9, p. 261-325, 1978.

BLOOMFIELD, Leonard. *Language*. Nova York: Holt, 1933.

BLOOR, David. *Knowledge and Social Imagery*. Londres: Routledge, 1976.

BLOUIN, Francis; ROSENBERG, William. *Processing the Past: Contesting Authority in History and the Archives*. Oxford: Oxford University Press, 2011.

BOAS, Franz. *Race, Language and Culture*. Chicago: University of Chicago Press, 1940.

BORGMAN, Christine. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge: MIT Press, 2015.

BORGMANN, Albert. *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University of Chicago Press, 1984.

BORODITSKY, Lera. Does Language Shape Thought? English and Mandarin Speakers' Conceptions of Time. *Cognitive Psychology*, v. 43, n. 1, p. 1-22, 2001.

BOSTROM, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

BOURDIEU, Pierre. *Ce que parler veut dire: l'économie des échanges linguistiques*. Paris: Fayard, 1982.

BOURDIEU, Pierre. *Outline of a Theory of Practice. Tradução de Richard Nice*. Cambridge: Cambridge University Press, 1977.

BOURDIEU, Pierre; DARBEL, Alain. *The Love of Art: European Art Museums and Their Public*. Cambridge: Polity Press, 1969/1991.

BOYLE, James. *The Public Domain: Enclosing the Commons of the Mind*. New Haven: Yale University Press, 2008.

BRANDI, Cesare. *Teoria da Restauração*. Cotia: Ateliê Editorial, 1963/2004.

BRATMAN, Michael E. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press, 2014.

BRENTANO, Franz. *Psychology from an Empirical Standpoint*. London: Routledge, 1874/1995.

BRIET, Suzanne. *Qu'est-ce que la documentation?* Paris: Éditions Documentaires Industrielles et Techniques, 1951.

BROOKS, Rodney. Intelligence without Representation. *Artificial Intelligence*, v. 47, n. 1-3, p. 139-159, 1991.

BROUWER, Luitzen Egbertus Jan. Intuitionism and Formalism. In: BENACERRAF, P.; PUTNAM, H. (eds.). *Philosophy of Mathematics: Selected Readings*. Cambridge: Cambridge University Press, 1913/1983. p. 77-89.

BROWN, Donald E. *Human Universals*. Nova York: McGraw-Hill, 1991.

BRUNER, Edward M. *Culture on Tour: Ethnographies of Travel*. Chicago: University of Chicago Press, 2005.

BUCKLAND, Michael K. Information as Thing. *Journal of the American Society for Information Science*, v. 42, n. 5, p. 351-360, 1991.

BULLER, David J. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge: MIT Press, 2005.

BUNGE, Mario. Technology as Applied Science. *Technology and Culture*, v. 7, n. 3, p. 329-349, 1966.

BURGE, Tyler. Content Preservation. *The Philosophical Review*, v. 102, n. 4, p. 457-488, 1993.

BUTERIN, Vitalik. The Meaning of Decentralization. *Medium*, 2017. Disponível em: https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274. Acesso em: 15 fev. 2026.

BYBEE, Joan. *Language, Usage and Cognition*. Cambridge: Cambridge University Press, 2010.

CABEZAS, Amalia L. *Economies of Desire: Sex and Tourism in Cuba and the Dominican Republic*. Philadelphia: Temple University Press, 2009.

CAPPELEN, Herman. *Philosophy without Intuitions*. Oxford: Oxford University Press, 2012.

CARNAP, Rudolf. *Der logische Aufbau der Welt*. Tradução de Rolf George. Chicago: Open Court, 1928/2003.

CARNAP, Rudolf. Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie*, v. 4, n. 11, p. 20-40, 1950.

CARNAP, Rudolf. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 1950.

CARR, David. *The Promise of Cultural Institutions*. Walnut Creek: AltaMira Press, 2003.

CARROLL, Lewis. What the Tortoise Said to Achilles. *Mind*, v. 4, n. 14, p. 278-280, 1895.

CASSIRER, Ernst. *Filosofia das Formas Simbólicas, vol. 1: A Linguagem*. São Paulo: Martins Fontes, 1923/1996.

CHALMERS, David. Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, v. 2, n. 3, p. 200-219, 1995.

CHALMERS, David. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.

CHOMSKY, Noam. *Aspects of the Theory of Syntax*. Cambridge: MIT Press, 1965.

CHOMSKY, Noam. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger, 1986.

CHOMSKY, Noam. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.

CHOULDECHOVA, Alexandra. *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*. Big Data, v. 5, n. 2, p. 153-163, 2017.

CHURCH, Alonzo. An Unsolvable Problem of Elementary Number Theory. *American Journal of Mathematics*, v. 58, n. 2, p. 345-363, 1936.

CHURCHLAND, Paul. Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*, v. 78, n. 2, p. 67-90, 1981.

CID, R. R. L. *Fundamentality: structures, powers, and a supervenience dualism*. Manuscrito. Disponível em: https://philpapers.org/archive/CIDFSP.pdf.

CID, R. R. L. O paradoxo inferencial de Lewis Carroll. *Fundamento*, v. 1, n. 12, 2017. https://doi.org/10.5281/v.j.v1i12.2437.

CID, R. R. L. *Refutações: ensaios em política, economia, ética e arte*. Porto Alegre: Fi, 2020.

CID, R. R. L. Uma visão de mundo filosófica. *Trans/Form/Ação*, v. 46, n. esp. 1, p. 571-596, 2023. https://doi.org/10.1590/0101-3173.2023.v46esp1.p571.

CLARK, Andy. *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press, 1997.

CLARK, Andy; CHALMERS, David. The Extended Mind. *Analysis*, v. 58, n. 1, p. 7-19, 1998.

CLIFFORD, James. R*outes: Travel and Translation in the Late Twentieth Century*. Cambridge: Harvard University Press, 1997.

COADY, C. A. J. *Testimony: A Philosophical Study*. Oxford: Oxford University Press, 1992.

COECKELBERGH, Mark. Robot Rights? Towards a Social-Relational Justification of Moral Consideration. *Ethics and Information Technology*, v. 12, n. 3, p. 209-221, 2010.

COHEN, Julie E. *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. New Haven: Yale University Press, 2012.

COHEN, Paul. The Independence of the Continuum Hypothesis. *Proceedings of the National Academy of Sciences*, v. 50, n. 6, p. 1143-1148, 1963.

COLBURN, Timothy R. *Philosophy and Computer Science*. Armonk: M. E. Sharpe, 2000.

COLLINS, Harry. *Tacit and Explicit Knowledge*. Chicago: University of Chicago Press, 2010.

COLLINS, Harry; PINCH, Trevor. *The Golem: What Everyone Should Know about Science*. Cambridge: Cambridge University Press, 1993.

COOK, Terry. Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era. *Archives and Manuscripts*, v. 22, n. 2, p. 300-328, 1994.

COOK, Terry. Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms. *Archival Science*, v. 13, n. 2-3, p. 95-120, 2013.

COOK, Terry. The Concept of the Archival Fonds in the Post-Custodial Era: Theory, Problems and Solutions. *Archivaria*, n. 35, p. 24-37, 1992.

CRANE, Diana. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago: University of Chicago Press, 1972.

CRESSWELL, Tim. *Place: A Short Introduction*. Oxford: Blackwell, 2004.

CRICK, Malcolm. Representations of International Tourism in the Social Sciences: Sun, Sex, Sights, Savings, and Servility. *Annual Review of Anthropology*, v. 18, p. 307-344, 1989.

CULLER, Jonathan. *Framing the Sign: Criticism and Its Institutions*. Norman: University of Oklahoma Press, 1988.

CULLER, Jonathan. Semiotics of Tourism. *American Journal of Semiotics*, v. 1, n. 1-2, p. 127-140, 1981.

CUNO, James. *Who Owns Antiquity? Museums and the Battle over Our Ancient Heritage*. Princeton: Princeton University Press, 2008.

CZAYKOWSKA-HIGGINS, Ewa. Research Models, Community Engagement, and Linguistic Fieldwork. *Language Documentation & Conservation*, v. 3, n. 1, p. 15-50, 2009.

DANIELS, Norman. Normal Functioning and the Treatment-Enhancement Distinction. *Cambridge Quarterly of Healthcare Ethics*, v. 9, n. 3, p. 309-322, 2000.

DAVIDSON, Donald. On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, v. 47, p. 5-20, 1973.

DAVIDSON, Donald. Radical Interpretation. *Dialectica*, v. 27, n. 1-4, p. 313-328, 1973.

DAVIS, Michael. *Thinking Like an Engineer: Studies in the Ethics of a Profession*. Oxford: Oxford University Press, 1998.

DAWKINS, Richard. *The Selfish Gene*. Oxford: Oxford University Press, 1976.

DE FILIPPI, Primavera; WRIGHT, Aaron. *Blockchain and the Law: The Rule of Code*. Cambridge: Harvard University Press, 2018.

DE MILLO, Richard A.; LIPTON, Richard J.; PERLIS, Alan J. Social Processes and Proofs of Theorems and Programs. *Communications of the ACM*, v. 22, n. 5, p. 271-280, 1979.

DEBORD, Guy. *A Sociedade do Espetáculo*. Rio de Janeiro: Contraponto, 1967/1997.

DENNETT, Daniel. *Consciousness Explained*. Boston: Little, Brown, 1991.

DERRIDA, Jacques. *De l'hospitalité*. Paris: Calmann-Lévy, 1997.

DERRIDA, Jacques. *Mal de arquivo: uma impressão freudiana*. Rio de Janeiro: Relume Dumará, 2001.

DESCARTES, René. *Discurso do Método*. Tradução de Maria Ermantina Galvão. São Paulo: Martins Fontes, 1637/2001.

DESCARTES, René. *Meditações sobre a filosofia primeira*. Tradução de Fausto Castilho. Campinas: Editora da Unicamp, 2004.

DEUTSCH, David. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society of London A*, v. 400, n. 1818, p. 97-117, 1985.

DEWEY, John. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan, 1916.

DEWEY, John. *Logic: The Theory of Inquiry*. Carbondale: Southern Illinois University Press, 1938/1986.

DEWEY, Melvil. *Dewey Decimal Classification and Relative Index*. 23. ed. Dublin: OCLC, 2011.

DIFFIE, Whitfield; HELLMAN, Martin. New Directions in Cryptography. *IEEE Transactions on Information Theory*, v. 22, n. 6, p. 644-654, 1976.

DIFFIE, Whitfield; LANDAU, Susan. *Privacy on the Line: The Politics of Wiretapping and Encryption*. Cambridge: MIT Press, 2007.

DIJKSTRA, Edsger W. Notes on Structured Programming. In: DAHL, Ole-Johan; DIJKSTRA, Edsger W.; HOARE, C. A. R. *Structured Programming*. London: Academic Press, 1972. p. 1-82.

DIJKSTRA, Edsger W. The Humble Programmer. *Communications of the ACM*, v. 15, n. 10, p. 859-866, 1972.

DILTHEY, Wilhelm. *Introdução às ciências humanas*. Tradução de Marco Casanova. Rio de Janeiro: Forense Universitária, 1883/2010.

DIXON, Chris. *Why Decentralization Matters. Medium*, 2018. Disponível em: https://medium.com/@cdixon/why-decentralization-matters-5e3f79f7638e. Acesso em: 15 fev. 2026.

DORIAN, Nancy C. A Response to Ladefoged's Other View of Endangered Languages. *Language*, v. 69, n. 3, p. 575-579, 1993.

DORST, Kees; CROSS, Nigel. Creativity in the Design Process: Co-evolution of Problem-Solution. *Design Studies*, v. 22, n. 5, p. 425-437, 2001.

DRETSKE, Fred I. *Knowledge and the Flow of Information*. Cambridge: MIT Press, 1981.

DUHEM, Pierre. *La théorie physique: son objet et sa structure*. Paris: Vrin, 1906/1991.

DUNCAN, Carol. *Civilizing Rituals: Inside Public Art Museums*. London: Routledge, 1995.

DURANTI, Luciana. *Diplomatics: New Uses for an Old Science*. Archivaria, n. 28, p. 7-27, 1998.

DURANTI, Luciana; MACNEIL, Heather. The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project. *Archivaria*, n. 42, p. 46-67, 1996.

DURANTI, Luciana; THIBODEAU, Kenneth. The Concept of a Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES. *Archival Science*, v. 6, n. 1, p. 13-68, 2006.

DURKHEIM, Émile. *As regras do método sociológico*. Tradução de Paulo Neves. São Paulo: Martins Fontes, 1895/2007.

EU, União Europeia. *Regulamento Geral sobre a Proteção de Dados (GDPR)*. Bruxelas: Parlamento Europeu, 2016.

EVANS, Nicholas. *Dying Words: Endangered Languages and What They Have to Tell Us*. Oxford: Wiley-Blackwell, 2010.

EVANS, Nicholas; LEVINSON, Stephen C. The Myth of Language Universals: Language Diversity and Its Importance for Cognitive Science. *Behavioral and Brain Sciences*, v. 32, n. 5, p. 429-448, 2009.

EVANS-PRITCHARD, E. E. *Witchcraft, Oracles and Magic among the Azande*. Oxford: Clarendon Press, 1937.

EVERETT, Daniel L. Cultural Constraints on Grammar and Cognition in Pirahã. *Current Anthropology*, v. 46, n. 4, p. 621-646, 2005.

FAIRCLOUGH, Norman. *Language and Power*. Londres: Longman, 1989.

FEENBERG, Andrew. *Transforming Technology: A Critical Theory Revisited*. Oxford: Oxford University Press, 2002.

FEINBERG, Joel. *The Moral Limits of the Criminal Law: Harm to Others*. Oxford: Oxford University Press, 1984.

FETZER, James H. *Program Verification: The Very Idea*. Communications of the ACM, v. 31, n. 9, p. 1048-1063, 1988.

FEYERABEND, Paul. Against Method: Outline of an Anarchistic Theory of Knowledge. London: New Left Books, 1975.

FIELD, Hartry. *Science Without Numbers*. Princeton: Princeton University Press, 1980.

FLORIDI, Luciano. *Information Ethics*: On the Philosophical Foundation of Computer Ethics. *Ethics and Information Technology*, v. 1, n. 1, p. 33-52, 1999.

FLORIDI, Luciano. *The Philosophy of Information*. Oxford: Oxford University Press, 2011.

FLORIDI, Luciano; COWLS, Josh. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, v. 1, n. 1, 2019.

FODOR, Jerry A. *Psychological Explanation*: An Introduction to the Philosophy of Psychology. New York: Random House, 1968.

FODOR, Jerry A. Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, v. 28, n. 2, p. 97-115, 1974.

FOUCAULT, Michel. *Surveiller et punir*: naissance de la prison. Paris: Gallimard, 1975.

FOUCAULT, Michel. *The Birth of the Clinic:* An Archaeology of Medical Perception. New York: Vintage, 1963/2003.

FRANKFURT, Harry G. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy,* v. 68, n. 1, p. 5-20, 1971.

FRASER, Nancy. From Redistribution to Recognition? Dilemmas of Justice in a "Post-Socialist" Age. *New Left Review*, n. 212, p. 68-93, 1995.

FREGE, Gottlob. On Sense and Reference [Über Sinn und Bedeutung]. In: GEACH, P.; BLACK, M. (eds.). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 1892/1980.

FREGE, Gottlob. *Os fundamentos da aritmética*. Tradução de Luís Henrique dos Santos. São Paulo: Abril Cultural, 1983. (Os Pensadores).

FREY, Carl Benedikt; OSBORNE, Michael A. The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, v. 114, p. 254-280, 2017.

FRICKÉ, Martin. The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, v. 35, n. 2, p. 131-142, 2009.

FRICKER, Miranda. *Epistemic injustice: power and the ethics of knowing*. Oxford: Oxford University Press, 2007.

GADAMER, Hans-Georg. *Verdade e método*. Tradução de Flávio Paulo Meurer. 10. ed. Petrópolis: Vozes, 1960/2008.

GARBER, Daniel. Understanding interaction: what Descartes should have told Elisabeth. *Southern Journal of Philosophy*, v. 21, supl., p. 15-32, 1983.

GAZDAR, Gerald et al. *Generalized phrase structure grammar*. Oxford: Blackwell, 1985.

GEERTZ, Clifford. *The interpretation of cultures*. New York: Basic Books, 1973.

GETTIER, Edmund L. Is justified true belief knowledge? *Analysis*, v. 23, n. 6, p. 121-123, 1963.

GIBSON, James J. *The ecological approach to visual perception*. New York: Psychology Press, 1979/2015.

GILBERT, Margaret. *Joint commitment: how we make the social world*. Oxford: Oxford University Press, 2013.

GILBERT, Margaret. Modelling collective belief. *Synthese*, v. 73, n. 1, p. 185-204, 1987.

GÖDEL, Kurt. O que é o problema do contínuo de Cantor? In: *Um texto inédito*. Lisboa: Fragmentos, 1989.

GÖDEL, Kurt. On formally undecidable propositions of *Principia Mathematica* and related systems. In: VAN HEIJENOORT, Jean (ed.). *From Frege to Gödel: a source book in mathematical logic*. Cambridge: Harvard University Press, 1931/1967. p. 592-616.

GOFFMAN, Erving. *The presentation of self in everyday life*. New York: Anchor Books, 1959.

GOLDMAN, Alvin I. *Knowledge in a social world*. Oxford: Oxford University Press, 1999.

GOLDREICH, Oded. *Foundations of cryptography: volume 1, basic tools*. Cambridge: Cambridge University Press, 2001.

GOLDWASSER, Shafi; MICALI, Silvio; RACKOFF, Charles. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, v. 18, n. 1, p. 186-208, 1989.

GOODMAN, Nelson. *Fact, fiction, and forecast*. Cambridge: Harvard University Press, 1955.

GOODMAN, Nelson. *Languages of art: an approach to a theory of symbols*. Indianapolis: Hackett, 1976.

GOODMAN, Nelson. *Ways of worldmaking*. Indianapolis: Hackett, 1978.

GÖSSLING, Stefan. *Tourism and global environmental change*. London: Routledge, 2006.

GÖSSLING, Stefan; COHEN, Scott. Why sustainable transport policies will fail: EU climate policy in the light of transport taboos. *Journal of Transport Geography*, v. 39, p. 196-207, 2014.

GRABURN, Nelson H. H. Tourism: the sacred journey. In: SMITH, Valene L. (ed.). *Hosts and guests: the anthropology of tourism*. Philadelphia: University of Pennsylvania Press, 1989. p. 21-36.

GRANGER, Stewart. Emulation as a digital preservation strategy. *D-Lib Magazine*, v. 6, n. 10, 2000.

GRETZEL, Ulrike; SIGALA, Marianna; XIANG, Zheng; KOO, Chulmo. Smart tourism: foundations and developments. *Electronic Markets*, v. 25, n. 3, p. 179-188, 2015.

GUREVICH, Yuri. Sequential abstract state machines capture sequential algorithms. *ACM Transactions on Computational Logic*, v. 1, n. 1, p. 77-111, 2000.

HABERMAS, Jürgen. *Moral consciousness and communicative action*. Tradução de Christian Lenhardt; Shierry Weber Nicholsen. Cambridge: MIT Press, 1990.

HABERMAS, Jürgen. *Teoria do agir comunicativo*. Tradução de Paulo Astor Soethe; Flávio Beno Siebeneichler. São Paulo: WMF Martins Fontes, 1981/2012.

HANSON, Norwood Russell. *Patterns of discovery*. Cambridge: Cambridge University Press, 1958.

HARDING, Sandra. *Whose science? whose knowledge? thinking from women's lives*. Ithaca: Cornell University Press, 1991.

HARDWIG, John. Epistemic dependence. *The Journal of Philosophy*, v. 82, n. 7, p. 335-349, 1985.

HARDY, Godfrey Harold. *A mathematician's apology*. Cambridge: Cambridge University Press, 1940/1992.

HARNAD, Stevan. The symbol grounding problem. *Physica D*, v. 42, p. 335-346, 1990.

HARRIS, Marvin. *Cultural materialism: the struggle for a science of culture*. New York: Random House, 1979.

HAUSER, Marc D.; CHOMSKY, Noam; FITCH, W. Tecumseh. The faculty of language: what is it, who has it, and how did it evolve? *Science*, v. 298, n. 5598, p. 1569-1579, 2002.

HAYLES, N. Katherine. *Writing machines*. Cambridge: MIT Press, 2002.

HEIDEGGER, Martin. Building dwelling thinking. In: *Poetry, language, thought*. New York: Harper Perennial, 1954/2008. p. 141-160.

HEIDEGGER, Martin. *Ser e tempo*. Petrópolis: Vozes, 1927/2015.

HEIN, George E. *Learning in the museum*. London: Routledge, 1998.

HENRICH, Joseph; HEINE, Steven J.; NORENZAYAN, Ara. The weirdest people in the world? *Behavioral and Brain Sciences*, v. 33, n. 2-3, p. 61-83, 2010.

HERSKOVITS, Melville J. *Cultural relativism: perspectives in cultural pluralism*. New York: Vintage Books, 1972.

HERSKOVITS, Melville J. *Man and his works: the science of cultural anthropology*. New York: Knopf, 1948.

HILBERT, David. On the infinite. In: VAN HEIJENOORT, Jean (ed.). *From Frege to Gödel: a source book in mathematical logic*. Cambridge: Harvard University Press, 1926/1967. p. 367-392.

HINTON, Leanne; HALE, Ken (orgs.). *The green book of language revitalization in practice*. San Diego: Academic Press, 2001.

HJØRLAND, Birger; ALBRECHTSEN, Hanne. Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, v. 46, n. 6, p. 400-425, 1995.

HOARE, C. A. R. An axiomatic basis for computer programming. *Communications of the ACM*, v. 12, n. 10, p. 576-580, 1969.

HOBSBAWM, Eric; RANGER, Terence (eds.). *The invention of tradition*. Cambridge: Cambridge University Press, 1983.

HOFSTADTER, Douglas. *Gödel, Escher, Bach: an eternal golden braid*. New York: Basic Books, 1979.

HOOPER-GREENHILL, Eilean. *Museums and the shaping of knowledge*. London: Routledge, 1992.

HORTON, Robin. African traditional thought and Western science. *Africa: Journal of the International African Institute*, v. 37, n. 1, p. 50-71; n. 2, p. 155-187, 1967.

HOUKES, Wybo; VERMAAS, Pieter E. *Technical functions: on the use and design of artefacts*. Dordrecht: Springer, 2010.

HUI, Yuk. *On the existence of digital objects*. Minneapolis: University of Minnesota Press, 2016.

HUME, David. *A treatise of human nature*. Oxford: Oxford University Press, 1739/2000.

HUME, David. *An enquiry concerning human understanding*. Oxford: Oxford University Press, 1748/1999.

HUSSERL, Edmund. *Ideias para uma fenomenologia pura e para uma filosofia fenomenológica*. Tradução de Márcio Suzuki. Aparecida: Ideias & Letras, 1913/2006.

INGARDEN, Roman. *The work of music and the problem of its identity*. Berkeley: University of California Press, 1931/1989.

JACKSON, Frank. Epiphenomenal qualia. *Philosophical Quarterly*, v. 32, n. 127, p. 127-136, 1982.

JACKSON, Frank. What Mary didn't know. *Journal of Philosophy*, v. 83, n. 5, p. 291-295, 1986.

JAMES, William. *The principles of psychology*. v. 1. New York: Henry Holt, 1890.

JASANOFF, Sheila (ed.). *States of knowledge: the co-production of science and social order*. London: Routledge, 2004.

JAUSS, Hans Robert. *A história da literatura como provocação à teoria literária*. São Paulo: Ática, 1974/1994.

JENKINS, Tiffany. *Keeping their marbles: how the treasures of the past ended up in museums and why they should stay there*. Oxford: Oxford University Press, 2016.

JENKINSON, Hilary. *A manual of archive administration*. London: Percy Lund, Humphries & Co., 1922.

JONAS, Hans. *O princípio responsabilidade: ensaio de uma ética para a civilização tecnológica*. Tradução de Marijane Lisboa; Luiz Barros Montez. Rio de Janeiro: Contraponto, 1979/2006.

KAHNEMAN, Daniel. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux, 2011.

KANT, Immanuel. *Crítica da razão pura*. Tradução de Manuela Pinto dos Santos; Alexandre Fradique Morujão. Lisboa: Fundação Calouste Gulbenkian, 2001.

KANT, Immanuel. *Fundamentação da metafísica dos costumes*. Tradução de Guido Antônio de Almeida. São Paulo: Discurso Editorial; Barcarolla, 1785/2005.

KETELAAR, Eric. Tacit narratives: the meanings of archives. *Archival Science*, v. 1, n. 2, p. 131-141, 2001.

KIRSHENBLATT-GIMBLETT, Barbara. *Destination culture: tourism, museums, and heritage*. Berkeley: University of California Press, 1998.

KITCHER, Philip. The division of cognitive labor. *Journal of Philosophy*, v. 87, n. 1, p. 5-22, 1990.

KLEINBERG, Jon; MULLAINATHAN, Sendhil; RAGHAVAN, Manish. Inherent trade-offs in the fair determination of risk scores. In: *Innovations in Theoretical Computer Science Conference*, 8., 2017. Proceedings… New York: ACM, 2017. Article 43.

KLEINBERG, Jon et al. Human decisions and machine predictions. *Quarterly Journal of Economics*, v. 133, n. 1, p. 237-293, 2016.

KNOBE, Joshua. Intentional action and side effects in ordinary language. *Analysis*, v. 63, n. 3, p. 190-194, 2003.

KNOBE, Joshua; NICHOLS, Shaun. An experimental philosophy manifesto. In: KNOBE, Joshua; NICHOLS, Shaun (eds.). *Experimental philosophy*. New York: Oxford University Press, 2008. p. 3-14.

KOBLITZ, Neal; MENEZES, Alfred. Another look at provable security. *Journal of Cryptology*, v. 20, n. 1, p. 3-37, 2007.

KOLMOGOROV, Andrey N. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, v. 1, n. 1, p. 1-7, 1965.

KRIPKE, Saul. *Naming and necessity*. Cambridge: Harvard University Press, 1980.

KROES, Peter; MEIJERS, Anthonie. The dual nature of technical artefacts. *Studies in History and Philosophy of Science*, v. 37, n. 1, p. 1-4, 2006.

KUHN, Thomas S. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.

KUHN, Thomas S. Objectivity, value judgment, and theory choice. In: KUHN, Thomas S. *The essential tension*. Chicago: University of Chicago Press, 1977. p. 320-339.

LABOV, William. *Principles of linguistic change: social factors*. Oxford: Blackwell, 2001.

LABOV, William. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LADYMAN, James; FRENCH, Steven. Science as structure: devising a new realist framework. *Philosophy of Science*, v. 70, n. 5, p. 1209-1219, 2003.

LAKATOS, Imre. Falsification and the methodology of scientific research programmes. In: LAKATOS, Imre; MUSGRAVE, Alan (eds.). *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press, 1970. p. 91-196.

LAKATOS, Imre. *The methodology of scientific research programmes: philosophical papers*. Cambridge: Cambridge University Press, 1978.

LANDAUER, Rolf. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, v. 5, n. 3, p. 183-191, 1961.

LARSEN, Jonas. Tourism mobilities and the travel glance. *Scandinavian Journal of Hospitality and Tourism*, v. 1, n. 2, p. 80-98, 2001.

LATOUR, Bruno. *Reassembling the social: an introduction to actor-network-theory*. Oxford: Oxford University Press, 2005.

LATOUR, Bruno. *Science in action: how to follow scientists and engineers through society*. Cambridge: Harvard University Press, 1987.

LATOUR, Bruno. Where are the missing masses? The sociology of a few mundane artifacts. In: BIJKER, W.; LAW, J. (eds.). *Shaping technology/building society*. Cambridge: MIT Press, 1992. p. 225-258.

LAUDAN, Larry. A confutation of convergent realism. *Philosophy of Science*, v. 48, n. 1, p. 19-49, 1981.

LAUDAN, Larry. *Progress and its problems*. Berkeley: University of California Press, 1977.

LAUDAN, Larry. The demise of the demarcation problem. In: COHEN, R. S.; LAUDAN, L. (eds.). *Physics, philosophy and psychoanalysis*. Dordrecht: Reidel, 1983. p. 111-127.

LAVOIE, Brian; DEMPSEY, Lorcan. Thirteen ways of looking at... digital preservation. *D-Lib Magazine*, v. 10, n. 7/8, 2004.

LEFEBVRE, Henri. *Le droit à la ville*. Paris: Anthropos, 1968.

LENZEN, Manfred et al. The carbon footprint of global tourism. *Nature Climate Change*, v. 8, p. 522-528, 2018.

LESSIG, Lawrence. *Code and other laws of cyberspace*. New York: Basic Books, 1999.

LEVESON, Nancy G. *Engineering a safer world: systems thinking applied to safety*. Cambridge: MIT Press, 2011.

LEVINE, Joseph. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, v. 64, n. 4, p. 354-361, 1983.

LEWIS, David. An argument for the identity theory. *Journal of Philosophy*, v. 63, n. 1, p. 17-25, 1966.

LEWIS, David. What experience teaches. In: LYCAN, William (ed.). *Mind and cognition*. Oxford: Blackwell, 1990. p. 499-519.

LIST, Christian; PETTIT, Philip. *Group agency: the possibility, design, and status of corporate agents*. Oxford: Oxford University Press, 2011.

LIPTON, Zachary C. The mythos of model interpretability. *Communications of the ACM*, v. 61, n. 10, p. 36-43, 2018.

LOAR, Brian. Phenomenal states. *Philosophical Perspectives*, v. 4, p. 81-108, 1990.

LONGINO, Helen. *The fate of knowledge*. Princeton: Princeton University Press, 2002.

LOW, Philip et al. *The Cambridge declaration on consciousness*. Cambridge, 2012. Disponível em: fcmconference.org.

MABILLON, Jean. *De re diplomatica*. Paris, 1681.

MACCANNELL, Dean. *The tourist: a new theory of the leisure class*. New York: Schocken Books, 1976.

MACHERY, Edouard. *Philosophy within its proper bounds*. Oxford: Oxford University Press, 2017.

MACHERY, Edouard; MALLON, Ron; NICHOLS, Shaun; STICH, Stephen P. Semantics, cross-cultural style. *Cognition*, v. 92, n. 3, p. B1-B12, 2004.

MADDY, Penelope. *Naturalism in mathematics*. Oxford: Oxford University Press, 1997.

MADDY, Penelope. *Realism in mathematics*. Oxford: Oxford University Press, 1990.

MARCUS, Gary; DAVIS, Ernest. *Rebooting AI: building artificial intelligence we can trust*. New York: Pantheon, 2019.

MARR, David. *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman, 1982.

MARX, Karl. *Manuscritos econômico-filosóficos*. São Paulo: Boitempo, 1844/2004.

MARX, Karl. *O capital: crítica da economia política*. v. 1. São Paulo: Nova Cultural, 1867/1996.

MASSEY, Doreen. *For space*. London: Sage, 2005.

MATTHIAS, Andreas. The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, v. 6, n. 3, p. 175-183, 2004.

MAYNARD SMITH, John. The concept of information in biology. *Philosophy of Science*, v. 67, n. 2, p. 177-194, 2000.

MCGINN, Colin. Can we solve the mind-body problem? *Mind*, v. 98, n. 391, p. 349-366, 1989.

MCMAHAN, Jeff. *The ethics of killing: problems at the margins of life*. Oxford: Oxford University Press, 2002.

MICHELS, Robert. *Political parties: a sociological study of the oligarchical tendencies of modern democracy*. New Brunswick: Transaction Publishers, 1911/1999.

MILANO, Claudio; CHEER, Joseph M.; NOVELLI, Marina (eds.). *Overtourism: excesses, discontents and measures in travel and tourism*. Wallingford: CABI, 2019.

MILL, John Stuart. *A system of logic, ratiocinative and inductive*. London: John W. Parker, 1843.

MILL, John Stuart. *An examination of Sir William Hamilton's philosophy*. Collected works, v. IX. Toronto: University of Toronto Press, 1865/1979.

MILL, John Stuart. *On liberty*. London: John W. Parker and Son, 1859.

MILLER, Seumas. *The moral foundations of social institutions: a philosophical study*. New York: Cambridge University Press, 2010.

MILLER, Tim. Explanation in artificial intelligence: insights from the social sciences. *Artificial Intelligence*, v. 267, p. 1-38, 2019.

MILROY, James; MILROY, Lesley. *Authority in language: investigating language prescription and standardisation*. London: Routledge, 1985.

MITCHAM, Carl. *Thinking through technology: the path between engineering and philosophy*. Chicago: University of Chicago Press, 1994.

MITCHELL, Tom M. The need for biases in learning generalizations. Technical Report CBM-TR-117. Rutgers University, 1980.

MUMFORD, Lewis. *Technics and civilization*. Chicago: University of Chicago Press, 1934/2010.

NAGEL, Thomas. What is it like to be a bat? *Philosophical Review*, v. 83, n. 4, p. 435-450, 1974.

NAKAMOTO, Satoshi. *Bitcoin: a peer-to-peer electronic cash system*. 2008. Disponível em: https://bitcoin.org/bitcoin.pdf. Acesso em: 10 mar. 2026.

NETTLE, Daniel; ROMAINE, Suzanne. *Vanishing voices: the extinction of the world's languages*. Oxford: Oxford University Press, 2000.

NEURATH, Otto; CARNAP, Rudolf; MORRIS, Charles (orgs.). *International encyclopedia of unified science*. Chicago: University of Chicago Press, 1938.

NICHOLS, Shaun; KNOBE, Joshua. Moral responsibility and determinism: the cognitive science of folk intuitions. *Noûs*, v. 41, n. 4, p. 663-685, 2007.

NISBETT, Richard E.; WILSON, Timothy D. Telling more than we can know: verbal reports on mental processes. *Psychological Review*, v. 84, n. 3, p. 231-259, 1977.

NISSENBAUM, Helen. Accountability in a computerized society. *Science and Engineering Ethics*, v. 2, n. 1, p. 25-42, 1996.

NISSENBAUM, Helen. *Privacy in context: technology, policy, and the integrity of social life*. Stanford: Stanford University Press, 2010.

NOZICK, Robert. *Anarchy, state, and utopia*. New York: Basic Books, 1974.

NUSSBAUM, Martha C. *Creating capabilities: the human development approach*. Cambridge: Harvard University Press, 2011.

NUSSBAUM, Martha C. *Women and human development: the capabilities approach*. Cambridge: Cambridge University Press, 2000.

O'NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown, 2016.

O'REILLY, Tim. What is Web 2.0: design patterns and business models for the next generation of software. O'Reilly Media, 2005.

OLSON, Hope. *The power to name: locating the limits of subject representation in libraries*. Dordrecht: Kluwer, 2002.

ONU. *Declaração universal dos direitos humanos*. Nova York: Nações Unidas, 1948.

ONU. *Declaração das Nações Unidas sobre os direitos dos povos indígenas*. Nova York: Nações Unidas, 2007.

ORTNER, Sherry B. Theory in anthropology since the sixties. *Comparative Studies in Society and History*, v. 26, n. 1, p. 126-166, 1984.

OSTROM, Elinor. *Governing the commons: the evolution of institutions for collective action*. Cambridge: Cambridge University Press, 1990.

OTLET, Paul. *Traité de documentation: le livre sur le livre*. Bruxelles: Éditions Mundaneum, 1934.

PARISER, Eli. *The filter bubble: what the internet is hiding from you*. New York: Penguin Press, 2011.

PASQUALE, Frank. *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.

PEARCE, Susan M. *Museums, objects and collections: a cultural study*. Washington: Smithsonian Institution Press, 1992.

PEIRCE, Charles Sanders. *Collected papers of Charles Sanders Peirce*. v. 1-6. Cambridge: Harvard University Press, 1931-1935.

PERROW, Charles. *Normal accidents: living with high-risk technologies*. New York: Basic Books, 1984.

PERZANOWSKI, Aaron; SCHULTZ, Jason. *The end of ownership: personal property in the digital economy*. Cambridge: MIT Press, 2016.

PETROSKI, Henry. *To engineer is human: the role of failure in successful design*. New York: St. Martin's Press, 1985.

PFEIFER, Rolf; BONGARD, Josh. *How the body shapes the way we think: a new view of intelligence*. Cambridge: MIT Press, 2006.

PINKER, Steven. *The blank slate: the modern denial of human nature*. New York: Viking, 2002.

PIZARRO, David. O que a filosofia experimental significa para a psicologia tradicional? *Filosofia Experimental | X-Phi*, 2009. Disponível em: https://filosofiaexperimental.wordpress.com. Acesso em: 15 fev. 2026.

PLACE, Ullin T. Is consciousness a brain process? *British Journal of Psychology*, v. 47, n. 1, p. 44-50, 1956.

PLANTINGA, Alvin. *Warrant and proper function*. Oxford: Oxford University Press, 1993.

PLATÃO. *A república*. Tradução de Maria Helena da Rocha Pereira. Lisboa: Fundação Calouste Gulbenkian, 2001.

POLANYI, Michael. *The tacit dimension*. London: Routledge & Kegan Paul, 1966.

POMIAN, Krzysztof. Colecção. In: *Enciclopédia Einaudi*. v. 1: Memória-História. Lisboa: Imprensa Nacional-Casa da Moeda, 1984. p. 51-86.

POPPER, Karl R. *Conjectures and refutations*. London: Routledge, 1963.

POPPER, Karl R. *The logic of scientific discovery*. London: Hutchinson, 1959.

POSTAL, Paul M. Remarks on the foundations of linguistics. *Philosophical Forum*, v. 34, n. 3-4, p. 233-252, 2003.

PRESTON, Beth. Why is a wing like a spoon? a pluralist theory of function. *Journal of Philosophy*, v. 95, n. 5, p. 215-254, 1998.

PRINZ, Jesse. Sexo em julgamento: mulheres e homens têm diferentes valores morais? *Filosofia Experimental | X-Phi*, 2010. Disponível em: https://filosofiaexperimental.wordpress.com. Acesso em: 15 fev. 2026.

PUTNAM, Hilary. *Mathematics, matter and method: philosophical papers*. v. 1. Cambridge: Cambridge University Press, 1975.

PUTNAM, Hilary. *Reason, truth and history*. Cambridge: Cambridge University Press, 1981.

PUTNAM, Hilary. *Realism with a human face*. Cambridge: Harvard University Press, 1990.

PUTNAM, Hilary. *Representation and reality*. Cambridge: MIT Press, 1988.

QUINE, Willard Van Orman. *Ontological relativity and other essays*. New York: Columbia University Press, 1969.

QUINE, Willard Van Orman. Two dogmas of empiricism. *Philosophical Review*, v. 60, n. 1, p. 20-43, 1951.

QUINE, Willard Van Orman. *Word and object*. Cambridge: MIT Press, 1960.

QUINE, Willard Van Orman; ULLIAN, Joseph S. *The web of belief*. New York: Random House, 1970.

RABINOW, Paul. *Reflections on fieldwork in Morocco*. Berkeley: University of California Press, 1977.

RABINOW, Paul. Representations are social facts: modernity and post-modernity in anthropology. In: CLIFFORD, James; MARCUS, George (eds.). *Writing culture*. Berkeley: University of California Press, 1986. p. 234-261.

RANGANATHAN, Shiyali Ramamrita. *The five laws of library science*. Madras: Madras Library Association, 1931.

RAWLS, John. *A theory of justice*. Cambridge: Harvard University Press, 1971.

RAZ, Joseph. *The morality of freedom*. Oxford: Clarendon Press, 1986.

REGAN, Tom. *The case for animal rights*. Berkeley: University of California Press, 1983.

RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTRIN, Carlos. Why should I trust you? explaining the predictions of any classifier. In: *Proceedings of KDD 2016*. New York: ACM, 2016. p. 1135-1144.

RICŒUR, Paul. *Do texto à ação: ensaios de hermenêutica II*. Tradução de Alcino Cartaxo; Maria José Sarabando. Porto: Rés, 1986/2008.

RICŒUR, Paul. *Interpretation theory: discourse and the surplus of meaning*. Fort Worth: Texas Christian University Press, 1976.

RIZZOLATTI, Giacomo; CRAIGHERO, Laila. The mirror-neuron system. *Annual Review of Neuroscience*, v. 27, p. 169-192, 2004.

ROGAWAY, Philip. The moral character of cryptographic work. *Cryptology ePrint Archive*, Report 2015/1162, 2015.

RORTY, Richard. *Philosophy and the mirror of nature*. Princeton: Princeton University Press, 1979.

SAHLINS, Marshall. *Culture and practical reason*. Chicago: University of Chicago Press, 1976.

SAID, Edward. *Orientalism*. New York: Pantheon Books, 1978.

SALMON, Wesley. *The foundations of scientific inference*. Pittsburgh: University of Pittsburgh Press, 1967.

SANDEL, Michael. *The case against perfection: ethics in the age of genetic engineering*. Cambridge: Harvard University Press, 2007.

SANTONI DE SIO, Filippo; VAN DEN HOVEN, Jeroen. Meaningful human control over autonomous systems: a philosophical account. *Frontiers in Robotics and AI*, v. 5, article 15, 2018.

SAPIR, Edward. The status of linguistics as a science. *Language*, v. 5, n. 4, p. 207-214, 1929.

SATIJA, M. P. *The theory and practice of the Dewey decimal classification system*. 2. ed. Oxford: Chandos, 2013.

SAUSSURE, Ferdinand de. *Curso de linguística geral*. Tradução de Antônio Chelini; José Paulo Paes; Izidoro Blikstein. São Paulo: Cultrix, 1916/2006.

SAYERS, W. C. Berwick. *An introduction to library classification*. London: Grafton, 1926.

SCHEYVENS, Regina. *Tourism for development: empowering communities*. Harlow: Prentice Hall, 2002.

SCHEYVENS, Regina. *Tourism and poverty*. London: Routledge, 2011.

SCHLICK, Moritz. Positivismus und realismus. *Erkenntnis*, v. 3, p. 1-31, 1932.

SCHWITZGEBEL, Eric. The unreliability of naive introspection. *Philosophical Review*, v. 117, n. 2, p. 245-273, 2008.

SCHWITZGEBEL, Eric; GARZA, Mara. A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, v. 39, n. 1, p. 98-119, 2015.

SEARLE, John R. Minds, brains, and programs. *Behavioral and Brain Sciences*, v. 3, n. 3, p. 417-424, 1980.

SEARLE, John R. *The construction of social reality*. New York: Free Press, 1995.

SEARLE, John R. *The rediscovery of the mind*. Cambridge: MIT Press, 1992.

SELLARS, Wilfrid. Philosophy and the scientific image of man. In: COLODNY, Robert (ed.). *Frontiers of science and philosophy*. Pittsburgh: University of Pittsburgh Press, 1962. p. 35-78.

SEN, Amartya. *Development as freedom*. New York: Knopf, 1999.

SEN, Amartya. Elements of a theory of human rights. *Philosophy and Public Affairs*, v. 32, n. 4, p. 315-356, 2004.

SHANNON, Claude E. Communication theory of secrecy systems. *Bell System Technical Journal*, v. 28, n. 4, p. 656-715, 1949.

SHANNON, Claude E.; WEAVER, Warren. *The mathematical theory of communication*. Urbana: University of Illinois Press, 1949.

SHAPIRO, Stewart. Conservativeness and incompleteness. *Journal of Philosophy*, v. 80, n. 9, p. 521-544, 1983.

SHAPIRO, Stewart. *Philosophy of mathematics: structure and ontology*. Oxford: Oxford University Press, 1997.

SHAPIRO, Stewart. *Thinking about mathematics: the philosophy of mathematics*. Oxford: Oxford University Press, 2000.

SHERA, Jesse H. *Sociological foundations of librarianship*. Bombay: Asia Publishing House, 1970.

SHOR, Peter W. Algorithms for quantum computation: discrete logarithms and factoring. In: *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*. Los Alamitos: IEEE, 1994. p. 124-134.

SHRADER-FRECHETTE, Kristin. *Risk and rationality: philosophical foundations for populist reforms*. Berkeley: University of California Press, 1991.

SIEG, Wilfried. On computability. In: IRVINE, A. D. (org.). *Philosophy of mathematics*. Amsterdam: Elsevier, 2009. p. 535-630.

SIMON, Herbert A. *The sciences of the artificial*. 3. ed. Cambridge: MIT Press, 1969/1996.

SIMPSON, Moira G. *Making representations: museums in the post-colonial era*. London: Routledge, 1996.

SINGER, Peter. *Animal liberation*. New York: HarperCollins, 1975.

SMART, J. J. C. Sensations and brain processes. *Philosophical Review*, v. 68, n. 2, p. 141-156, 1959.

SMITH, Linda Tuhiwai. *Decolonizing methodologies: research and indigenous peoples*. London: Zed Books, 1999.

SONTAG, Susan. *On photography*. New York: Farrar, Straus and Giroux, 1977.

SOSA, Ernest. *A virtue epistemology: apt belief and reflective knowledge*. v. 1. Oxford: Oxford University Press, 2007.

SPOLSKY, Joel. The law of leaky abstractions. Joel on Software, 2002. Disponível em: https://www.joelonsoftware.com/2002/11/11/the-law-of-leaky-abstractions/.

SPIVAK, Gayatri Chakravorty. Can the subaltern speak? In: NELSON, Cary; GROSSBERG, Lawrence (eds.). *Marxism and the interpretation of culture*. Urbana: University of Illinois Press, 1988. p. 271-313.

SPROUSE, Jon; ALMEIDA, Diogo. Assessing the reliability of textbook data in syntax: Adger's core syntax. *Journal of Linguistics*, v. 48, n. 3, p. 609-652, 2012.

STICH, Stephen P. What every speaker knows. *Philosophical Review*, v. 80, n. 4, p. 476-496, 1971.

STOLER, Ann Laura. *Along the archival grain: epistemic anxieties and colonial common sense*. Princeton: Princeton University Press, 2009.

STROUD, Barry. *Hume*. London: Routledge, 1977.

THIBODEAU, Kenneth. Overview of technological approaches to digital preservation. In: *The state of digital preservation: an international perspective*. Washington: Council on Library and Information Resources, 2002. p. 4-31.

THOMPSON, Dennis F. Moral responsibility of public officials: the problem of many hands. *American Political Science Review*, v. 74, n. 4, p. 905-916, 1980.

TOMASELLO, Michael. *Constructing a language: a usage-based theory of language acquisition*. Cambridge: Harvard University Press, 2003.

TONONI, Giulio. An information integration theory of consciousness. *BMC Neuroscience*, v. 5, n. 42, 2004.

TONONI, Giulio. Consciousness as integrated information: a provisional manifesto. *Biological Bulletin*, v. 215, n. 3, p. 216-242, 2008.

TOOBY, John; COSMIDES, Leda. The psychological foundations of culture. In: BARKOW, Jerome H.; COSMIDES, Leda; TOOBY, John (eds.). *The adapted mind*. Oxford: Oxford University Press, 1992. p. 19-136.

TUAN, Yi-Fu. *Space and place: the perspective of experience*. Minneapolis: University of Minnesota Press, 1977.

TURING, Alan M. Computing machinery and intelligence. *Mind*, v. 59, n. 236, p. 433-460, 1950.

TURING, Alan M. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s. 2, v. 42, n. 1, p. 230-265, 1936.

TURKLE, Sherry. *Alone together: why we expect more from technology and less from each other*. New York: Basic Books, 2011.

TURNER, Raymond. *Computational artifacts: towards a philosophy of computer science*. Berlin: Springer, 2018.

TURNER, Stephen. *Liberal democracy 3.0*. London: Sage, 2003.

TURNER, Victor. *The ritual process: structure and anti-structure*. Chicago: Aldine, 1969.

URRY, John. *The tourist gaze*. London: Sage, 1990.

VALLOR, Shannon. Care ethics and human-robot interaction. *Ethics and Information Technology*, v. 13, n. 3, p. 207-218, 2011.

VALLOR, Shannon. *Technology and the virtues*. Oxford: Oxford University Press, 2016.

VAN DIJK, Jan. *The network society*. 2. ed. London: Sage, 2006.

VAN FRAASSEN, Bas C. *The scientific image*. Oxford: Oxford University Press, 1980.

VAN GELDER, Timothy. What might cognition be, if not computation? *Journal of Philosophy*, v. 92, n. 7, p. 345-381, 1995.

VARELA, Francisco; THOMPSON, Evan; ROSCH, Eleanor. *The embodied mind*. Cambridge: MIT Press, 1991.

VARDI, Moshe Y. What is an algorithm? *Communications of the ACM*, v. 55, n. 3, p. 5, 2012.

VEBLEN, Thorstein. *The theory of the leisure class*. New York: Penguin, 1899/1994.

VINCENTI, Walter G. *What engineers know and how they know it*. Baltimore: Johns Hopkins University Press, 1990.

WALZER, Michael. *Thick and thin*. Notre Dame: University of Notre Dame Press, 1994.

WANG, Ning. Rethinking authenticity in tourism experience. *Annals of Tourism Research*, v. 26, n. 2, p. 349-370, 1999.

WEINBERG, Jonathan; NICHOLS, Shaun; STICH, Stephen. Normativity and epistemic intuitions. *Philosophical Topics*, v. 29, n. 1-2, p. 429-460, 2001.

WEINREICH, Max. YIVO un di problemen fun undzer tsayt. *YIVO Bleter*, v. 25, n. 1, p. 3-18, 1945.

WEIZENBAUM, Joseph. *Computer power and human reason*. San Francisco: W. H. Freeman, 1976.

WESTIN, Alan F. *Privacy and freedom*. New York: Atheneum, 1967.

WHEELER, John A. Information, physics, quantum. In: ZUREK, Wojciech H. (ed.). *Complexity, entropy and the physics of information*. Redwood City: Addison-Wesley, 1990. p. 3-28.

WHORF, Benjamin Lee. *Language, thought and reality*. Cambridge: MIT Press, 1956.

WIENER, Norbert. *The human use of human beings*. Boston: Houghton Mifflin, 1950.

WIGNER, Eugene. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics*, v. 13, n. 1, p. 1-14, 1960.

WILSON, Edward O. *Sociobiology*. Cambridge: Harvard University Press, 1975.

WINCH, Peter. *The idea of a social science and its relation to philosophy*. London: Routledge, 1958.

WINCH, Peter. Understanding a primitive society. *American Philosophical Quarterly*, v. 1, n. 4, p. 307-324, 1964.

WINNER, Langdon. Do artifacts have politics? *Daedalus*, v. 109, n. 1, p. 121-136, 1980.

WITTGENSTEIN, Ludwig. *Investigações filosóficas*. Tradução de Marcos G. Montagnoli. Petrópolis: Vozes, 1953/2009.

WODAK, Ruth; MEYER, Michael (orgs.). *Methods of critical discourse analysis*. 2. ed. London: Sage, 2009.

WONG, David. *Natural moralities*. Oxford: Oxford University Press, 2006.

WORRALL, John. Structural realism: the best of both worlds? *Dialectica*, v. 43, n. 1-2, p. 99-124, 1989.

YUSTE, Rafael; GOERING, Sara. Four ethical priorities for neurotechnologies and AI. *Nature*, v. 551, p. 159-163, 2017.

ZAHAVI, Dan. Phenomenology and the project of naturalization. *Phenomenology and the Cognitive Sciences*, v. 3, n. 4, p. 331-347, 2004.

ZUBOFF, Shoshana. *The age of surveillance capitalism*. New York: PublicAffairs, 2019.

Tabelas

1. Filosofia das Ciências

Domínio	Problema	Teorias/Soluções	Argumentos Principais
Metafísica	Realismo científico	Realismo; antirrealismo; empirismo construtivo	Sucesso da ciência vs. subdeterminação
Metafísica	Natureza das leis	Regularismo; necessitarismo; disposicionalismo	Uniformidade vs. poderes causais
Epistemologia	Indução	Indutivismo; falsificacionismo; bayesianismo	Justificação vs. circularidade
Epistemologia	Demarcação	Verificacionismo; falsificacionismo; paradigma	Critérios vs. historicidade
Ética	Responsabilidade científica	Neutralidade axiológica vs. ciência engajada	Autonomia vs. impactos sociais

2. Filosofia da Antropologia

Domínio	Problema	Teorias/Soluções	Argumentos
Metafísica	Natureza da cultura	Idealismo cultural; materialismo cultural; sociobiologia	Significado vs. causalidade
Metafísica	Universalismo humano	Essencialismo vs. construtivismo	Natureza humana vs. variação
Epistemologia	Conhecimento intercultural	Relativismo; universalismo	Tradução de conceitos
Ética	Julgamento moral intercultural	Relativismo moral; universalismo moral	Tolerância vs. crítica

3. Filosofia da Linguística

Domínio	Problema	Teorias	Argumentos
Metafísica	O que é uma língua	Platonismo; mentalismo; externalismo social	Competência vs. uso
Epistemologia	Evidência linguística	Intuições; corpus; experimentação	Competência idealizada
Ética	Políticas linguísticas	Prescritivismo; descritivismo; justiça linguística	Poder e exclusão

4. Filosofia da Matemática

Domínio	Problema	Teorias	Argumentos
Metafísica	Existência dos objetos matemáticos	Platonismo; nominalismo; estruturalismo	Objetividade vs. ontologia inflada
Epistemologia	Conhecimento matemático	Apriorismo; empirismo; intuicionismo	Necessidade vs. experiência
Ética	Aplicação social da matemática	Neutralidade vs. responsabilidade	Modelos e decisões públicas

5. Filosofia das Ciências da Informação

Domínio	Problema	Teorias	Argumentos
Metafísica	Natureza da informação	Informação física; semântica; pragmática	Dados vs. significado
Epistemologia	Autoridade epistêmica	Veritismo; construtivismo social	Verdade vs. poder
Ética	Acesso à informação	Abertura vs. controle	Democracia informacional

6. Filosofia da Arquivologia

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Natureza do documento arquivístico; ontologia do registro; identidade e autenticidade	Realismo documental; teoria dos traços; ontologia social dos registros	Documento como evidência institucional ancorada em práticas sociais
Epistemologia	Valor probatório; verdade arquivística; memória vs. registro	Diplomática contemporânea; epistemologia da evidência	Confiabilidade baseada em proveniência e cadeia de custódia
Ética	Poder arquivístico; silenciamento; justiça informacional	Arquivos como instrumentos de poder (pós-estruturalismo)	Arquivística crítica e pluralização da memória

7. Filosofia da Biblioteconomia

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Natureza da obra, edição e exemplar	Ontologia bibliográfica; FRBR/LRM	Obra como entidade abstrata instanciada
Epistemologia	Organização do conhecimento; classificação vs. realidade	Teorias da indexação; epistemologia social	Sistemas híbridos e flexíveis
Ética	Acesso à informação; censura; neutralidade	Biblioteconomia crítica	Biblioteca como agente democrático

8. Filosofia da Museologia

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Estatuto do objeto museal; autenticidade	Teoria da aura; ontologia do artefato	Objeto como nó de relações culturais
Epistemologia	Narrativa expositiva; construção do passado	História pública; curadoria interpretativa	Museu como produtor de conhecimento
Ética	Colonialismo museal; repatriação	Museologia social	Descolonização e participação comunitária

9. Filosofia do Turismo

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Autenticidade da experiência turística	Teoria da encenação	Turismo como performance social
Epistemologia	Conhecimento do "outro" mediado	Estudos culturais	Reflexividade do turista
Ética	Impacto cultural e ambiental	Turismo sustentável	Responsabilidade global

10. Filosofia da Engenharia e da Tecnologia

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Natureza dos artefatos técnicos	Filosofia da técnica	Artefatos como mediadores do mundo
Epistemologia	Conhecimento projetual vs. científico	Epistemologia do design	Engenharia como saber prático
Ética	Risco tecnológico; responsabilidade	Ética da tecnologia	Princípio da precaução

11. Filosofia da Ciência da Computação

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Natureza da informação e do algoritmo	Realismo informacional	Computação como estrutura formal do mundo
Epistemologia	Prova, simulação e verificação	Epistemologia computacional	Simulação como forma legítima de conhecimento
Ética	Privacidade; vigilância; poder algorítmico	Ética da informação	Governança algorítmica

12. Filosofia da Criptografia

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Natureza do segredo e da confiança	Ontologia da informação protegida	Segurança como propriedade relacional
Epistemologia	Conhecimento sob sigilo; prova de conhecimento	Zero-knowledge proofs	Saber sem revelar
Ética	Privacidade vs. segurança estatal	Criptoanarquismo vs. soberania	Equilíbrio entre liberdade e controle

13. Filosofia da Ciência Cognitiva e da Mente

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Natureza da mente; consciência	Fisicalismo; funcionalismo	Mente como processo emergente
Epistemologia	Conhecimento interno vs. externo	Cognição incorporada	Conhecimento distribuído
Ética	Manipulação cognitiva; neuroética	Ética da mente	Proteção da autonomia mental

14. Filosofia da Robótica e da Inteligência Artificial

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Estatuto ontológico da IA	Artificialidade vs. agência	Sistemas artificiais como agentes limitados
Epistemologia	Inteligência artificial vs. humana	Teste de Turing; IA forte/fraca	Inteligência como desempenho funcional
Ética	Responsabilidade algorítmica; autonomia	Ética da IA	Princípios de alinhamento e controle

15. Filosofia das Artes Digitais

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Obra digital e reprodução infinita	Ontologia da arte digital	Obra como processo interativo
Epistemologia	Experiência estética mediada	Estética tecnológica	Novo regime perceptivo
Ética	Autoria, IA e direitos	Pós-autoria	Arte como criação híbrida humano-máquina

16. Filosofia da Filosofia Experimental

Dimensão	Problemas Filosóficos	Teorias / Argumentos	Soluções / Posições
Metafísica	Intuições como dados filosóficos	Naturalização da filosofia	Filosofia empiricamente informada
Epistemologia	Método experimental na filosofia	X-phi	Integração com ciências cognitivas
Ética	Uso de sujeitos e vieses culturais	Crítica metodológica	Universalização cautelosa

Sobre os Autores

Rodrigo Reis Lastra Cid é professor de filosofia na Universidade Federal de Ouro Preto (UFOP), onde atua na graduação e no Programa de Pós-Graduação em Filosofia. Fez doutorado e mestrado em Lógica e Metafísica pela UFRJ, com período sanduíche na Université Catholique de Louvain (Bélgica), e pós-doutorado pela UFMG. Antes de chegar a Ouro Preto, foi professor na Universidade Federal do Amapá (UNIFAP) e na UFRJ. É coordenador geral da Série Investigação Filosófica (NEPFIL/Editora UFPel), projeto que já traduziu e publicou dezenas de livros com apoio da John Templeton Foundation. Lidera o Grupo Investigação Filosófica (DGP/CNPq) e o laboratório GIFLABS, dedicado a explorar as interseções entre filosofia, blockchain e tecnologias educacionais. Sua pesquisa concentra-se em metafísica das leis da natureza, modalidade, filosofia da física e, mais recentemente, filosofia da tecnologia e da arte digital. Além da carreira acadêmica, atua como artista plástico e digital, com obras expostas em galerias no Brasil e no exterior.

Matheus dos Reis Gomes é doutorando em Filosofia pela UFOP, com bolsa do CNPq, onde investiga as relações entre fenomenologia, psicanálise e filosofia da mente. Mestre e bacharel em Filosofia pela UFJF, também licenciado em Filosofia pela mesma universidade, e especialista em Ciência da Religião. Sua trajetória acadêmica se estrutura a partir da articulação entre a fenomenologia husserliana, a psicanálise freudiana e a logoterapia de Viktor Frankl, articulação que se condensa em uma investigação sistemática dos conceitos de consciência, inconsciente e subjetividade. Tem-se dedicado também ao ensino de filosofia, à tradução de textos fundamentais e à reflexão sobre metodologias dialógicas em sala de aula. Atua como revisor de periódicos acadêmicos e colabora em projetos de pesquisa sobre ética, cognição e filosofia da psicanálise.

Este livro é fruto de uma parceria que começou nas aulas do doutorado, onde ambos se conheceram. Matheus foi tradutor e revisor em um dos volumes da Série Investigação Filosófica (*Textos Selecionados de Filosofia da Vida e da Morte*), e hoje trabalha ao lado de Rodrigo em outro projeto da série, dedicado à Filosofia Computacional, organizado por Rodrigo e por seu colega Pedro Luiz Caetano Filho. Essa colaboração contínua é parte de uma visão de ensino que

Rodrigo carrega desde que era aluno de graduação e que agora se desdobra em sua atuação como professor na Universidade Federal de Ouro Preto, a saber, a de que a filosofia se aprende fazendo, e que o lugar do estudante de pós-graduação não é apenas o de assimilar conhecimento, mas o de se tornar, ele mesmo, um produtor de filosofia.

O livro que o leitor tem em mãos é, de certo modo, um registro dessa aposta. Matheus não é apenas um aluno brilhante, mas um parceiro de escrita que, desde o começo, soube trazer suas próprias perguntas, suas leituras, sua maneira de pensar, para dentro do texto. Cada capítulo foi pensado e revisto a quatro mãos, num movimento de ida e volta que misturava discussões sobre ontologia e metafísica com conversas sobre fenomenologia, psicanálise e arte digital. Acreditamos que o resultado é um livro que reflete essa pluralidade de vozes, essa recusa em reduzir a filosofia a uma única tradição, esse entusiasmo por juntar o que às vezes se insiste em separar.

Escrevemos este livro porque acreditamos que a filosofia das ciências não é um exercício de especialistas trancados em suas torres de marfim, mas uma conversa que envolve todos os que se perguntam o que significa conhecer, preservar, criar e transformar o mundo. Que ele possa, para quem o ler, ser também um convite para pensar em conjunto.


Editora
UFPel


DISSERTATIO
FILOSOFIA