

Série Investigação Filosófica

Textos selecionados de  
**Filosofia da Cognição**

**Eros Moreira de Carvalho**  
(Organizador)

# **TEXTOS SELECIONADOS DE FILOSOFIA DA COGNIÇÃO**

Série Investigação Filosófica

**TEXTOS SELECIONADOS DE FILOSOFIA DA COGNIÇÃO**

Eros Moreira de Carvalho  
(Organizador)



Pelotas, 2022

## **REITORIA**

Reitora: Isabela Fernandes Andrade

Vice-Reitora: Ursula Rosa da Silva

Chefe de Gabinete: Aline Ribeiro Paliga

Pró-Reitor de Graduação: Maria de Fátima Cóssio

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

Pró-Reitor de Extensão e Cultura: Eraldo dos Santos Pinheiro

Pró-Reitor de Planejamento e Desenvolvimento: Paulo Roberto Ferreira Júnior

Pró-Reitor Administrativo: Ricardo Hartlebem Peter

Pró-Reitor de Gestão de Informação e Comunicação: Julio Carlos Balzano de Mattos

Pró-Reitor de Assuntos Estudantis: Fabiane Tejada da Silveira

Pró-Reitor de Gestão Pessoas: Taís Ulrich Fonseca

## **CONSELHO EDITORIAL DA EDITORA DA UFPEL**

Presidente do Conselho Editorial: Ana da Rosa Bandeira

Representantes das Ciências Agrônômicas: Victor Fernando Büttow Roll

Representantes da Área das Ciências Exatas e da Terra: Eder João Lenardão

Representantes da Área das Ciências Biológicas: Rosângela Ferreira Rodrigues

Representante da Área das Engenharias e Computação: Reginaldo da Nóbrega Tavares

Representantes da Área das Ciências da Saúde: Fernanda Capella Rugno

Representante da Área das Ciências Sociais Aplicadas: Daniel Lena Marchiori Neto

Representante da Área das Ciências Humanas: Charles Pereira Pennaforte

Representantes da Área das Linguagens e Artes: Lúcia Bergamaschi Costa Weymar

## **EDITORIA DA UFPEL**

Chefia: Ana da Rosa Bandeira (Editora-chefe)

Seção de Pré-produção: Isabel Cochrane (Administrativo)

Seção de Produção: Suelen Aires Böettge (Administrativo)

Anelise Heidrich (Revisão)

Ingrid Fabiola Gonçalves (Diagramação)

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

Morgana Riva (Assessoria)

### **CONSELHO EDITORIAL**

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

### **COMISSÃO TÉCNICA (EDITORAÇÃO)**

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

### **DIREÇÃO DO IFISP**

Prof. Dr. João Hobuss

### **CHEFE DO DEPARTAMENTO DE FILOSOFIA**

Prof. Dr. Juliano Santos do Carmo

## **Série Investigação Filosófica**

A Série Investigação Filosófica, uma iniciativa do **Núcleo de Ensino e Pesquisa em Filosofia** do Departamento de Filosofia da UFPel e do **Grupo de Pesquisa Investigação Filosófica** do Departamento de Filosofia da UNIFAP, sob o selo editorial do NEPFil online e da Editora da Universidade Federal de Pelotas, tem por objetivo precípuo a publicação da tradução para a língua portuguesa de textos selecionados a partir de diversas plataformas internacionalmente reconhecidas, tal como a *Stanford Encyclopedia of Philosophy* (<https://plato.stanford.edu/>), por exemplo. O objetivo geral da série é disponibilizar materiais bibliográficos relevantes tanto para a utilização enquanto material didático quanto para a própria investigação filosófica.

### **EDITORES DA SÉRIE**

Rodrigo Reis Lastra Cid (IF/UNIFAP)  
Juliano Santos do Carmo (NEPFIL/UFPEL)

### **COMISSÃO TÉCNICA**

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

### **ORGANIZADOR DO VOLUME**

Eros Moreira de Carvalho (UFRGS)

### **TRADUTORES E REVISORES**

Bruno Malavolta e Silva (UFRGS)  
Eros Moreira de Carvalho (UFRGS)  
Jeferson Diello Huffermann (UFRGS)  
Jessica Nunes Vergara (IFSUL)  
Pedro Maggi Rech Noguez (UFRGS)  
Rodrigo de Ulhôa Canto Reis (UFRGS)  
Thiago Gruner (UFRGS)  
César Fernando Meurer (UFABC)  
Filipe Lazzari (UFG)  
Nara M. Figueiredo (CLE-Unicamp)  
Sofia Stein (UNISINOS)

### **CRÉDITO DA IMAGEM DE CAPA**

RENI, Guido. Salome with the Head of Saint John the Baptist. Source:  
<https://www.artic.edu/artworks/11434/salome-with-the-head-of-saint-john-the-baptist>



## **GRUPO DE PESQUISA INVESTIGAÇÃO FILOSÓFICA (UNIFAP/CNPq)**

O Grupo de Pesquisa Investigação Filosófica (DPG/CNPq) foi constituído por pesquisadores que se interessam pela investigação filosófica nas mais diversas áreas de interesse filosófico. O grupo foi fundado em 2010, como grupo independente, e se oficializou como grupo de pesquisa da Universidade Federal do Amapá em 2019.

### **MEMBROS PERMANENTES DO GRUPO**

Aluizio de Araújo Couto Júnior  
Bruno Aislã Gonçalves dos Santos  
Cesar Augusto Mathias de Alencar  
Daniel Schiochett  
Daniela Moura Soares  
Everton Miguel Puhl Maciel  
Guilherme da Costa Assunção Cecílio  
Kherian Galvão Cesar Gracher  
Luiz Helvécio Marques Segundo  
Paulo Roberto Moraes de Mendonça  
Pedro Merlussi  
Rafael César Pitt  
Rafael Martins  
Renata Ramos da Silva  
Rodrigo Alexandre de Figueiredo  
Rodrigo Reis Lastra Cid  
Sagid Salles  
Tiago Luís Teixeira de Oliveira

## © Série Investigação Filosófica, 2022

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

Universidade Federal do Amapá  
Departamento de Filosofia  
Grupo de Pesquisa Investigação Filosófica

### NEPFil online

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

Os direitos autorais estão de acordo com a Política Editorial do NEPFil online. As revisões ortográficas e gramaticais foram realizadas pelos tradutores e revisores. A autorização para a tradução dos verbetes da *Stanford Encyclopedia of Philosophy* neste volume foi obtida pelo *Grupo de Pesquisa Investigação Filosófica*.

**Primeira publicação em 2022 por NEPFil online e Editora da UFPel.**

### Dados Internacionais de Catalogação

---

N123      Textos selecionados de filosofia da cognição.  
[recurso eletrônico] Organizador: Eros Moreira de Carvalho – Pelotas: NEPFIL Online,  
2022.  
424p. - (Série Investigação Filosófica).  
Modo de acesso: Internet  
<wp.ufpel.edu.br/nepfil>  
ISBN: 978-65-86440-87-9

1. Filosofia. 2. Cognição I. CARVALHO, Eros Moreira de.

COD 100

---



Para maiores informações, visite o site [wp.ufpel.edu.br/nepfil](http://wp.ufpel.edu.br/nepfil)



# SUMÁRIO

<b>Sobre a série investigação filosófica</b>	14
<b>Introdução</b>	16
<b>(I) Ciência Cognitiva</b>	20
1. História	21
2. Métodos	22
3. Representação e Computação	25
4. Abordagens Teóricas	26
4.1. Lógica Formal	26
4.2. Regras	27
4.3. Conceitos	27
4.4. Analogias	28
4.5. Imagens	29
4.6. Conexionismo	30
4.7. Neurociência Teórica	30
4.8. Bayesianismo	31
4.9. Aprendizagem Profunda	32
5. Relevância Filosófica	33
5.1. Aplicações Filosóficas	33
5.2. Crítica à Ciência Cognitiva	35
5.3. Filosofia da Ciência Cognitiva	36
Referência bibliográfica	37
<b>(II) A Teoria Computacional da Mente</b>	42
1. Máquinas de Turing	43
2. Inteligência artificial	46
3. A teoria clássica computacional da mente	48
3.1 Funcionalismo de máquina	51

3.2 A teoria representacional da mente	53
4. Redes neurais	56
4.1 Relações entre a computação das redes neurais e a computação clássica	59
4.2 Argumentos a favor do conexionismo	61
4.3 Sistemática e produtividade	64
4.4 Neurociência computacional	67
5. Computação e representação	70
5.1 Computação como formal	72
5.2 Externalismo acerca do conteúdo mental	76
5.3 Computação envolvente-de-conteúdo	80
6. Concepções alternativas de computação	83
6.1 Processamento de informação	84
6.2 Resolução de funções	86
6.3 Estruturalismo	88
6.4 Teorias mecanicistas	91
6.5 Pluralismo	92
7. Argumentos contra o computacionalismo	92
7.1 Argumentos de trivialidade	93
7.2 Teorema da incompletude do Gödel	94
7.3 Limites da modelagem computacional	94
7.4 Argumentos temporais	96
7.5 Cognição corporificada	97
Referência bibliográfica	98

### **(III) Teorias Teleológicas do Conteúdo Mental**

1. Objetivos gerais	112
2. Funções Teleológicas	118
3. Teorias Teleossemânticas	123
3.1 Semântica do Indicador	126
3.2 Teorias Baseadas em Benefícios e Consumidores	132
3.3 Teorias não-combinatórias	137
3.4 Teorias Combinatórias mais ou menos modestas	140
3.5 Teorias Causais-Informacionais	142
4. Problemas para a Teleossemântica	146
4.1 Indeterminação Funcional	146
4.2 Homem do Pântano	151

4.3 Conceitos e capacidades sofisticadas	157
Referência bibliográfica	159
<b>(IV) Modularidade da Mente</b>	166
1. O que é um módulo mental?	167
2. Modularidade, à la Fodor: Uma proposta modesta	173
2.1. Desafios à modularidade a nível básico	174
2.2. O argumento de Fodor contra a modularidade de alto nível	179
3. Modularidade pós-Fodor	183
3.1. O argumento em prol da modularidade massiva	185
3.2. Dúvidas sobre a modularidade massiva	189
4. Modularidade e a filosofia	191
Referência bibliográfica	192
<b>(V) Cognição Corporificada</b>	197
1. Ciência Cognitiva Tradicional vs. Ciência Cognitiva Corporificada	198
2. Algumas Bases Históricas da Ciência Cognitiva Corporificada	200
2.1 Metáfora e Cognição	201
2.2 Cognição Enativa	202
2.3 Repensando a Robótica	205
2.4 Percepção Ecológica	206
2.5 Dinamismo e Desenvolvimento	207
2.6 Fenomenologia	207
3. O que a Cognição Corporificada é	208
4. Corporificação vs. Tradição sobre Três Questões	211
4.1 Modularidade	213
4.2 Representação Mental	216
4.3 Nativismo	219
5. Domínios Empíricos para a Cognição Corporificada	224
5.1 Consciência Visual	224
5.2 Conceitos	228
5.3 Memória	230
5.4 Outras Mentes	231
5.5 Cognição Moral	233
6. Divisões mais nítidas na Cognição Corporificada	236
6.1 Ganhos para a pesquisa empírica	237

6.2 Acomodação pela ciência cognitiva tradicional	238
6.3 Cognição corporificada e a tese da mente estendida	239
6.4 Agência, self e subjetividade	241
Referência bibliográfica	243

## **(VI) Emoção** 261

1. Definindo as emoções: quais são os <i>desiderata</i> ?	262
2. Três tradições no estudo das emoções: emoções como sentimentos, avaliações e motivações	265
3. A tradição sentimentalista inicial: emoções como sentimentos	267
4. Emoções e objetos intencionais	269
5. A tradição avaliativa inicial: emoções como julgamentos	271
6. A tradição avaliativa na ciência dos afetos: teorias da apreciação	274
7. A tradição híbrida avaliativo-sentimental na filosofia recente	278
7.1 Emoções como percepções avaliativas	278
7.2 Emoções como sentimentos avaliativos	280
7.3 Emoções como padrões de saliência	282
8. A tradição motivacional na ciência dos afetos e seus oponentes	284
8.1 Teoria da emoção básica: emoções como programas de afeto evoluídos	285
8.2 A concepção ecológico-comportamental, o construtivismo psicológico e o construcionismo social: emoções como construções	287
9. A tradição motivacional na filosofia recente	292
9.1 As teorias motivacional e atitudinal: emoções como atitudes e estados-motivo	292
9.2 As teorias enativistas das emoções: emoções como enações	295
10. Racionalidade e emoções	299
10.1 Racionalidade cognitiva como adequação, justificação e coerência	299
10.2 Racionalidade estratégica instrumental e substantiva	301
11. Considerações finais	305
Referência bibliográfica	309

## **(VII) Cognição Animal** 327

1. O que é Cognição Animal?	328
2. Questões Fundacionais	329
2.1 Mentes Animais	330
2.2 Antropomorfismo e Psicologia do Senso Comum	332

2.3 Racionalidade	336
2.4 Crença	339
2.5 Conceitos	342
3. A Ciência da Cognição Animal	347
3.1 Método Anedótico Inicial	347
3.2 Métodos Experimentais	349
3.3 Métodos Observacionais	354
4. Programas de Pesquisa	357
4.1 Comunicação	358
4.1.1 Sinais Expressivos e Referenciais	358
4.1.2 Sinais Intencionais	361
4.1.3 Comunicação Simbólica	364
4.1.4 Comunicação Gestual	371
4.2 Leitura de Mentes ou Teoria da Mente	372
4.2.1 Leitura das Percepções da Mente [ <i>Mindreading Perceptions</i> ]	374
4.2.2 O Problema Lógico e a Parcimônia	376
4.2.3 Além dos Chimpanzés: Leitura de Mente em Outras Espécies	380
4.3 Autorreconhecimento em Espelho	382
4.4 Metacognição, Memória, e Monitoramento de Incerteza	383
4.5 Prática Moral	385
4.5.1 Emoção e Empatia	386
4.5.2 Cooperação	389
4.5.3 Punição	391
5. Cognição Animal e Filosofia: O Que Vem a Seguir?	393
Referência bibliográfica	393
<b>Sobre os organizadores</b>	<b>420</b>
<b>Sobre os tradutores e revisores</b>	<b>421</b>

# **SOBRE A SÉRIE INVESTIGAÇÃO FILOSÓFICA**

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

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

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

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

# INTRODUÇÃO

Após os excelentes volumes I e II de *Textos Selecionados de Filosofia da Ciência da Série Investigações Filosóficas*, que oferecem recursos e ferramentas fundamentais para pensar e refletir sobre a ciência e a prática científica em geral, podemos começar a nos dedicar, também, a questões mais específicas que surgem quando voltamos a atenção para uma ciência ou disciplina científica em particular. Sendo assim, o foco deste volume são as ciências cognitivas, um empreendimento científico que surgiu entre os anos 50 e 60, do século XX, e envolve a colaboração de disciplinas como a filosofia, a psicologia, a inteligência artificial, a antropologia, a neurociência e a linguística, com a finalidade de que se realizem o estudo e a investigação dos fenômenos cognitivos.

A filosofia das ciências cognitivas compreende tanto questões de segunda ordem, acerca da disciplina em questão, quanto questões de primeira ordem, que dizem respeito a tópicos relacionados à cognição. São questões do primeiro tipo: Como confirmamos ou refutamos um modelo cognitivo? Que tipo de explicação melhor explica a cognição? Podemos integrar os estudos das diferentes áreas que compõem as ciências cognitivas em uma teoria abrangente da cognição? Os diferentes níveis de explicação, neurológico, fisiológico, computacional, comportamental, etc., são melhor caracterizados em termos ontológicos ou metodológicos? Quais são os principais fundamentos e recursos teóricos e explicativos das ciências cognitivas? Já as questões do segundo tipo: O que é a cognição? Emoções envolvem alguma forma de cognição? A percepção é infiltrada por crenças e expectativas? Quais mecanismos e processos cognitivos dão suporte ao fenômeno da linguagem? Quais as relações entre mente, cérebro, corpo e ambiente? O que é o pensamento? Máquinas podem pensar? Muitas dessas questões são compartilhadas com a filosofia da mente ou com as próprias ciências cognitivas. O diferencial da filosofia das ciências cognitivas é o comprometimento na abordagem dessas questões, integrando as pesquisas realizadas nas ciências cognitivas em uma concepção ou teoria mais abrangente. Em alguns casos, até avançando novas hipóteses que



poderão vir a ser investigadas empiricamente pelas ciências cognitivas. Por vezes, a segunda bateria de questões é referida como *filosofia nas ciências cognitivas*.

Quando o professor Rodrigo Reis, organizador da série, procurou-me para participar dos volumes e sugeriu que eu organizasse um livro na área de filosofia da ciência, na época, eu ministrava uma disciplina sobre filosofia das ciências cognitivas na graduação. Apesar de já termos uma boa produção especializada nesta área no próprio país, uma dificuldade não negligenciável para montar esse curso foi encontrar material introdutório e abrangente em língua portuguesa. Sendo assim, ocorreu-me a ideia de organizar um volume sobre filosofia das ciências cognitivas, para o qual selecionei verbetes que permitem uma boa entrada ao projeto inicial das ciências cognitivas, centrado nas noções de representação e computação, bem como às críticas que esse projeto recebeu e continua recebendo a partir de reflexões sobre as relações entre corpo, afetividade e mente. Este livro apresenta os seguintes verbetes: (I) Ciência Cognitiva, (II) A Teoria Computacional da Mente, (III) Teorias Teleológicas do Conteúdo Mental, (IV) Modularidade da Mente, (V) Cognição Corporificada, (VI) Emoção e (VII) Cognição Animal.

O primeiro capítulo, *Ciência Cognitiva*, oferece uma breve introdução à disciplina, abrangendo tanto a sua formação quanto os seus métodos, principais pressupostos teóricos, padrões explicativos e objetos de estudo. No final, há uma discussão sobre a relevância filosófica da disciplina. Na sequência, trazemos um conjunto de capítulos que articula o projeto clássico das ciências cognitivas, segundo o qual a cognição é computação sobre representações. Esse projeto é normalmente chamado de *Cognitivismo Clássico*. O primeiro deles, *A Teoria Computacional da Mente*, apresenta as noções de computação, algoritmo e de máquina de Turing. Essas noções são centrais para explorar a metáfora da mente como um computador. Em que consistiria dizer que a mente computa? Podemos entender todos os processos mentais em termos computacionais? Além de discutir diferentes compreensões da computação, o capítulo também explora as limitações e dificuldades da teoria computacional da mente. Na sequência, passamos à discussão da noção de conteúdo mental ou representação. Se a mente pode ser vista como um computador, o que ela exatamente computa? Uma resposta é que ela computa ou processa representações. O que nos leva à pergunta de se essa noção pode ser naturalizada e fazer parte de uma teoria empírica da mente. Essa é a discussão central do capítulo *Teorias Teleológicas do Conteúdo Mental*. O terceiro capítulo desse conjunto dedicado à articulação do cognitivismo clássico é *Modularidade da Mente*. Qual é a arquitetura da mente? Se a mente pode ser entendida em termos

de funções computacionais, e essas funções podem ser descritas em termos de algoritmos, então tanto melhor que a arquitetura da mente tenha uma organização modular, em que cada módulo é dedicado a uma função computacional específica. Alegadamente, a seleção natural também corrobora a ideia de que a mente deve ter uma organização modular. Neste capítulo, discute-se como esses módulos podem ser compreendidos e se, enquanto recursos explicativos, eles são suficientes para acomodar e explicar os diversos tipos de cognição que realizamos. Finalizada a apresentação do cognitivismo clássico, passamos a um conjunto de capítulos que vão explorar aspectos mais específicos da mente e da cognição. O capítulo *Cognição Corporificada* condensa uma série de reflexões sobre a relação estreita entre mente e corpo que desde os anos 80 e 90 animam muitos debates nas ciências cognitivas, forçando a rearticulação do cognitivismo clássico ou mesmo fomentando a emergência de novas propostas de programas de pesquisa nas ciências cognitivas. O capítulo explora as várias maneiras pelas quais a mente pode ser entendida como estruturada e constituída pelo próprio corpo e como isso nos força a questionar a metáfora da mente como um computador ou a pelo menos reconhecer algumas limitações substantivas dessa metáfora para pensar a mente como um todo. No capítulo seguinte, *Emoção*, entramos na discussão de um fenômeno muito caro às ciências cognitivas mas que também esteve no horizonte da filosofia desde os seus primórdios. Qual a relação entre a afetividade e a cognição? As emoções têm objetos intencionais? Elas realizam ou embutem alguma espécie de avaliação? Elas são sensíveis a razões? O capítulo contempla diferentes abordagens e respostas a essas questões, tanto abordagens mais alinhadas ao cognitivismo clássico quanto abordagens mais alinhadas às ideias da mente corporificada. Por fim, no último capítulo, *Cognição Animal*, entramos no rico e cada vez mais explorado reino das mentes dos animais não-humanos. A investigação sobre essas mentes levanta questões metodológicas e teóricas substantivas. Ao mesmo tempo, ela enriquece as discussões das ciências cognitivas na medida em que alguns dos seus resultados surpreendentes nos levam a repensar a compreensão que temos de algumas capacidades e habilidades cognitivas. As discussões em torno da mente animal também permeiam e reverberam nas controvérsias entre o cognitivismo clássico e as ideias da mente corporificada.

Esta obra não teria sido possível sem o trabalho atento e primoroso de Bruno Malavolta e Silva, César Fernando Meurer, Filipe Lazzeri, Jeferson Diello Huffermann, Jessica Nunes Vergara, Nara M. Figueredo, Pedro Maggi Rech Noguez, Rodrigo de Ulhôa Canto Reis, Sofia Stein e Thiago Gruner. Sou muito grato pela participação de todos vocês neste projeto. O rico material que os leitores têm em

mãos se deve a esses tradutores e revisores, bem como aos organizadores gerais da *Série Investigações Filosóficas*, Rodrigo Reis Lastra Cid e Juliano do Carmo, além do Edward Zalta, da *Stanford Encyclopedia of Philosophy*, que tornaram esta iniciativa possível.

Eros Moreira de Carvalho (UFRGS/CNPq)  
Organizador do Livro

# (I) Ciência Cognitiva\*

Autor: Paul Thagard

Tradução: Eros Moreira de Carvalho (UFRGS)

Revisão: Sofia Stein (UNISINOS)

A ciência cognitiva<sup>1</sup> é o estudo interdisciplinar da mente e da inteligência, abrangendo a filosofia, a psicologia, a inteligência artificial, a neurociência, a linguística e a antropologia. Ela surgiu em meados dos anos 50 quando pesquisadores de várias áreas começaram a desenvolver teorias da mente com base em

---

\* THAGARD, T. Cognitive Science. In: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2018. Disponível em: <https://plato.stanford.edu/archives/win2018/entries/cognitive-science/>. Acesso em: 10 out. 2021.

The following is the translation of the entry on Cognitive Science by Paul Thagard, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/win2018/entries/cognitive-science/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/cognitive-science/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.

<sup>1</sup>N.T.: Não é incomum encontrar a expressão “ciências cognitivas”. No entanto, a diferença entre a expressão no plural e no singular é usada para marcar a diferença entre uma versão mais fraca da disciplina, em que ela é apenas o somatório de outras disciplinas, e uma versão mais forte, em que há convergência teórica e metodológica das diversas disciplinas que se juntam para investigar e compreender a cognição. O autor do verbete usa a expressão “ciência cognitiva” no singular para marcar a expectativa de uma convergência teórica na área e o acompanharemos assim na tradução.

representações complexas e procedimentos computacionais. A sua origem organizacional data de meados dos anos 70, quando a Sociedade da Ciência Cognitiva foi fundada e iniciou o periódico *Cognitive Science*. Desde então, mais de cem universidades na América do Norte, Europa, Ásia e Austrália constituíram programas de ciência cognitiva, e muitas outras criaram cursos de ciência cognitiva.

## 1. História

Tentativas de entender a mente e o seu funcionamento remontam pelo menos aos gregos antigos, quando filósofos como Platão e Aristóteles tentaram explicar a natureza do conhecimento humano. O estudo da mente permaneceu no domínio da filosofia até o século dezenove, quando a psicologia experimental se desenvolveu. Wilhem Wundt e seus estudantes inauguraram métodos de laboratório para estudar as operações mentais de modo mais sistemático. Em poucas décadas, contudo, a psicologia experimental foi dominada pelo behaviorismo, uma concepção que praticamente negava a existência da mente. De acordo com os behavioristas, como J. B. Watson, a psicologia deveria se restringir ao exame da relação entre estímulos observáveis e respostas comportamentais observáveis. O discurso sobre consciência e representações mentais foi banido da discussão científica considerada respeitável. Especialmente na América do Norte, o behaviorismo dominou a comunidade psicológica até os anos 50.

Por volta de 1956, a paisagem intelectual começou a mudar dramaticamente. George Miller resenhou inúmeros estudos que mostravam que a capacidade humana de pensar é limitada. A memória de curto prazo, por exemplo, seria limitada a aproximadamente sete itens. Ele propôs então que as limitações da memória podem ser superadas recodificando informações em blocos, representações mentais que requerem procedimentos para codificar e decodificar a informação. Nesta época, computadores primitivos eram encontrados há apenas alguns poucos anos, mas pioneiros como John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon estavam fundando a área de inteligência artificial. Além disso, Noam Chomsky rejeitou as suposições behavioristas sobre a linguagem como um hábito aprendido e propôs no lugar explicar a compreensão linguística em termos de gramáticas mentais constituídas por regras. Os seis pensadores mencionados neste parágrafo podem ser considerados os fundadores da ciência cognitiva.

## 2. Métodos

A ciência cognitiva contém ideias teóricas unificadoras, porém precisamos levar em consideração a diversidade de concepções e métodos que pesquisadores de diferentes áreas trazem para o estudo da mente e da inteligência. Embora os psicólogos cognitivistas hoje atuem frequentemente na teorização e na modelagem computacional, o seu método fundamental é a experimentação com seres humanos. Pessoas, geralmente estudantes de graduação que satisfazem certos requisitos, são trazidas para o laboratório de forma a que diferentes tipos de pensamento possam ser estudados sob condições controladas. Por exemplo, os psicólogos têm examinado experimentalmente os tipos de erros que as pessoas fazem durante o raciocínio dedutivo, as maneiras pelas quais as pessoas formam e aplicam conceitos, a velocidade das pessoas ao pensarem com imagens mentais, e o desempenho das pessoas ao resolverem problemas usando analogias. As conclusões sobre como a mente funciona devem ser baseadas em mais do que o “senso comum” e a introspecção, visto que esses podem dar uma imagem enganadora das operações mentais, muitas das quais não são acessíveis à consciência. Cada vez mais, os psicólogos selecionam os participantes, de fontes culturais diversas, para experimentos em plataformas de colaboração coletiva, como a *Amazon Mechanical Turk*<sup>2</sup>. Experimentos psicológicos que abordam cuidadosamente as operações mentais a partir de diferentes perspectivas são, portanto, cruciais para a ciência cognitiva ser científica. A experimentação é uma metodologia também empregada pela filosofia experimental.

Embora a teoria sem experimentação seja vazia, a experimentação sem teoria é cega. Para abordar as questões centrais sobre a natureza da mente, os experimentos psicológicos precisam ser interpretáveis em uma estrutura teórica que postula representações mentais e procedimentos. Uma das melhores maneiras de elaborar estruturas teóricas é pela formulação e teste de modelos computacionais que pretendem ser análogos às operações mentais. Para complementar os experimentos psicológicos sobre o raciocínio dedutivo, a formação de conceitos, a imaginação mental, e a solução de problemas por analogia, pesquisadores desenvolveram modelos computacionais que simulam aspectos das ações humanas.

---

<sup>2</sup>N.T.: Sobre a *Amazon Mechanical Turk*, confira o site, disponível em: <https://www.mturk.com/>. Acesso em: 05 out.

Projetar, construir e realizar experimentos com modelos computacionais é o método central da inteligência artificial (IA), o ramo da ciência da computação que se ocupa com sistemas inteligentes. Idealmente, na ciência cognitiva, os modelos computacionais e a experimentação psicológica andam de mãos dadas, mas muito trabalho importante em IA tem investigado o poder de diferentes abordagens da representação do conhecimento com relativa independência da psicologia experimental.

Embora alguns linguistas realizem experimentos psicológicos ou desenvolvam modelos computacionais, a maioria atualmente usa métodos diferentes. Para linguistas da tradição chomskiana, a principal tarefa teórica é a de identificar princípios gramaticais que forneçam a estrutura básica das linguagens humanas. Essa identificação ocorre atentando para diferenças sutis entre enunciações gramaticais e não gramaticais. Em português, por exemplo, as frases “Ela acertou a bola” e “Do que você gosta?” são gramaticais, mas “Ela a acertou bola” e “Do que você gostam?” não são. Uma gramática do português explicará por que as primeiras são aceitáveis mas não as últimas. Uma abordagem alternativa, a linguística cognitiva, coloca menos ênfase na sintaxe e mais na semântica e nos conceitos.

Assim como os psicólogos cognitivistas, os neurocientistas frequentemente realizam experimentos controlados, mas as suas observações são muito diferentes, visto que os neurocientistas estão preocupados precisamente com a natureza do cérebro. Os pesquisadores podem inserir eletrodos e gravar os disparos de neurônios individuais em animais não-humanos. Em relação aos seres humanos para os quais essa técnica seria demasiadamente invasiva, é agora comum usar dispositivos de escaneamento por ressonância magnética e por emissão de pósitrons para observar o que está acontecendo em diferentes partes do cérebro enquanto as pessoas realizam várias tarefas mentais. Por exemplo, os escaneamentos cerebrais identificaram as regiões do cérebro envolvidas na imaginação mental e na interpretação de palavras. Evidência adicional sobre o funcionamento cerebral é coletada observando a atuação de pessoas cujos cérebros foram danificados de formas identificáveis. Um derrame, por exemplo, em uma parte do cérebro dedicada à linguagem pode produzir deficiências, como a incapacidade de proferir frases. Tal qual a psicologia cognitiva, a neurociência é frequentemente tanto teórica quanto experimental, e o desenvolvimento da teoria é frequentemente auxiliado pelo desenvolvimento de modelos computacionais do comportamento de grupos de neurônios.

A antropologia cognitiva expande o exame do pensamento humano ao considerar como o pensamento funciona em diferentes contextos culturais. O estudo da mente não deveria obviamente se restringir a como os falantes do português

pensam, mas deveria considerar possíveis diferenças entre culturas nas maneiras de pensar. A ciência cognitiva está se tornando cada vez mais consciente da necessidade de ver as operações da mente em ambientes físicos e sociais particulares. Para antropólogos culturais, o principal método é a etnografia, que requer viver e interagir com membros de uma cultura o suficiente para que os seus sistemas sociais e cognitivos se tornem aparentes. Antropólogos cognitivos têm investigado, por exemplo, as similaridades e diferenças entre culturas nas palavras para designar cores.

Tradicionalmente, os filósofos não fazem observações empíricas sistemáticas ou constroem modelos computacionais, embora tenha havido um aumento recente de trabalhos na filosofia experimental. Contudo, a filosofia permanece importante para a ciência cognitiva porque ela lida com questões fundamentais que subjazem a abordagem experimental e computacional da mente. Questões abstratas como a natureza da representação e da computação não precisam ser investigadas na prática cotidiana da psicologia ou inteligência artificial, mas elas inevitavelmente surgem quando os pesquisadores refletem profundamente sobre o que estão fazendo. A filosofia também lida com questões gerais como a relação entre mente e corpo e com questões metodológicas como a natureza das explicações encontradas na ciência cognitiva. Além disso, a filosofia se preocupa com questões normativas sobre como as pessoas deveriam pensar, bem como com questões descritivas sobre como elas pensam. Além do objetivo teórico de entender o pensamento humano, a ciência cognitiva pode ter o objetivo prático de melhorá-lo, o que exige a reflexão normativa sobre o que queremos que o pensamento seja. A filosofia da mente não tem um método específico, porém ela deve compartilhar com os melhores trabalhos teóricos em outras áreas uma preocupação com resultados empíricos.

Na sua forma mais fraca, a ciência cognitiva é apenas a soma das disciplinas mencionadas: psicologia, inteligência artificial, linguística, neurociência, antropologia e filosofia. O trabalho interdisciplinar se torna muito mais interessante quando há convergência teórica e experimental acerca das conclusões sobre a natureza da mente. Por exemplo, a psicologia e a inteligência artificial podem ser combinadas por meio de modelos computacionais de como as pessoas se comportam em experimentos. A melhor maneira de compreender a complexidade do pensamento humano é usar múltiplos métodos, especialmente experimentos psicológicos e neurológicos e modelos computacionais. Teoricamente, a abordagem mais fecunda tem sido a de entender a mente em termos de representação e computação.



### 3. Representação e Computação

A hipótese central da ciência cognitiva é a de que o pensamento pode ser melhor compreendido em termos de estruturas representacionais na mente e procedimentos computacionais que operam nessas representações. Embora haja muito desacordo sobre a natureza das representações e computações que constituem o pensamento, a hipótese central é suficientemente geral para abranger o escopo atual da noção de pensamento na ciência cognitiva, incluindo as teorias conexionistas que modelam o pensamento usando redes neuronais artificiais.

A maior parte da pesquisa na ciência cognitiva supõe que a mente tem representações mentais análogas às estruturas de dados do computador, e procedimentos computacionais similares aos algoritmos computacionais. Os teóricos cognitivistas têm proposto que a mente contém essas representações mentais na forma de proposições lógicas, regras, conceitos, imagens e analogias, e que usa procedimentos mentais tais como dedução, busca, equiparação, rotação e recuperação. A analogia dominante mente-computador na ciência cognitiva tomou uma nova direção com o uso de outro análogo, o cérebro.

Os conexionistas propuseram ideias novas sobre a representação e a computação que usam neurônios e as suas conexões como inspirações para estruturas de dados, e disparos neuronais e a difusão da ativação como inspirações para algoritmos. A ciência da cognição trabalha, pois, com uma analogia complexa de três vias entre a mente, o cérebro e computadores. Cada um deles, a mente, o cérebro ou a computação, pode ser usado para sugerir novas ideias acerca dos outros. Não há um único modelo computacional da mente, visto que diferentes tipos de computadores e abordagens de programação sugerem diferentes maneiras de como a mente funciona. Os computadores que a maioria de nós usa atualmente são processadores seriais, realizando uma instrução por vez, mas o cérebro e alguns computadores construídos mais recentemente são processadores paralelos, capazes de fazer várias operações simultaneamente.

Uma tendência importante na ciência cognitiva atual é a integração da neurociência com muitas áreas da psicologia, incluindo a psicologia cognitiva, social, do desenvolvimento e clínica. Essa integração é parcialmente experimental, e resulta de uma explosão de novos instrumentos de estudo do cérebro, tais como a ressonância magnética funcional, a estimulação magnética transcraniana e a optogenética. A integração também é teórica, por causa dos avanços na compreensão de como grandes populações de neurônios podem realizar tarefas que são normalmente explicadas por teorias cognitivas acerca de regras e conceitos.

## 4. Abordagens Teóricas

Aqui está um resumo esquemático das teorias atuais sobre a natureza das representações e computações que explicam como a mente funciona.

### 4.1 Lógica Formal

A lógica formal fornece algumas ferramentas poderosas para examinar a natureza da representação e da computação. Os cálculos proposicional e de predicados servem para expressar muitos tipos complexos de conhecimento e muitas inferências podem ser compreendidas em termos de dedução lógica com regras inferenciais tais como o *modus ponens*. O esquema explicativo para a abordagem lógica é:

*Objetivo da explicação:*

- Por que as pessoas fazem as inferências que elas fazem?

*Padrão explicativo:*

- As pessoas têm representações mentais similares às frases do cálculo de predicados.
- As pessoas usam procedimentos dedutivos e indutivos que operam sobre essas frases.
- Os procedimentos dedutivos e indutivos, aplicados às frases, produzem as inferências.

Não é certo, contudo, que a lógica forneça as ideias centrais sobre a representação e a computação necessárias para a ciência cognitiva, dado que métodos de computação mais eficientes e psicologicamente naturais podem ser necessários para explicar o pensamento humano.

## 4.2 Regras

Muito do conhecimento humano é naturalmente descrito em termos de regras da forma **se** [...] **então** [...], e muitos tipos de pensamento, tal como o planejamento, podem ser modelados por sistemas baseados em regras. O esquema explicativo usado é:

*Objetivo da explicação:*

- Por que as pessoas têm um tipo particular de comportamento inteligente?

*Padrão explicativo:*

- As pessoas têm regras mentais.
- As pessoas têm procedimentos de uso dessas regras para procurar em um conjunto de possíveis soluções, e procedimentos para gerar novas regras.
- Os procedimentos para usar e formar regras produzem o comportamento.

Os modelos computacionais baseados em regras forneceram simulações detalhadas de uma ampla gama de experimentos psicológicos, desde a resolução de problemas criptoaritméticos até a aquisição de habilidades e o uso da linguagem. Os sistemas baseados em regras também têm sido de importância prática ao sugerir como melhorar a aprendizagem e como desenvolver sistemas de máquinas inteligentes.

## 4.3 Conceitos

Os conceitos, que parcialmente correspondem às palavras na linguagem falada e escrita, são um tipo importante de representação mental. Há razões psicológicas e computacionais para abandonar a concepção clássica de que conceitos têm definições estritas. Em vez disso, conceitos podem ser vistos como conjuntos de características típicas. A aplicação de conceitos é então uma questão de obter uma equiparação aproximada entre conceitos e o mundo. Esquemas e rotinas são mais complexos que conceitos que correspondem a palavras, mas eles são similares por consistirem em agrupamentos de características que podem ser equiparados e aplicados a novas situações. O esquema explicativo usado em sistemas baseados em conceitos é:

*Objetivo da explicação:*

- Por que as pessoas têm um tipo particular de comportamento inteligente?

*Padrão explicativo:*

- As pessoas têm um conjunto de conceitos, organizados por meio de compartimentos que estabelecem hierarquias de tipo e parte e outras associações.
- As pessoas têm um conjunto de procedimentos para a aplicação de conceitos, incluindo difusão de ativação, coordenação e sequência.
- Os procedimentos aplicados aos conceitos produzem o comportamento.
- Conceitos podem ser traduzidos em regras, mas eles agrupam informação de modo diferente de conjuntos de regras, tornando possível procedimentos computacionais diferentes.

## 4.4 Analogias

As analogias desempenham um papel importante no pensamento humano, em áreas tão diversas como a solução de problemas, a tomada de decisão, a explicação e a comunicação linguística. Modelos computacionais simulam como as pessoas recuperam e mapeiam análogos para aplicá-los às situações alvo. O esquema explicativo para analogias é:

*Objetivo da explicação:*

- Por que as pessoas têm um tipo particular de comportamento inteligente?

*Padrão explicativo:*

- As pessoas têm representações verbais e visuais de situações que podem ser usadas como casos ou análogos.
- As pessoas têm procedimentos de recuperação, mapeamento e adaptação que operam sobre análogos.
- Os processos analógicos, aplicados a representações de análogos, produzem o comportamento.

As restrições de similaridade, estrutura e propósito superam o problema difícil de como experiências prévias podem ser encontradas e usadas para ajudar com novos problemas. Nem todo o pensamento é analógico, e usar analogias inapropriadas pode impedir o pensamento, mas as analogias podem ser muito efetivas em aplicações tais como na educação e na elaboração de projetos.

## 4.5 Imagens

Imagens visuais ou de outros tipos desempenham um papel importante no pensamento humano. Representações pictóricas capturam a informação visual e espacial em um formato muito mais útil do que longas descrições verbais. Procedimentos computacionais bem ajustados para representações visuais incluem inspeção, busca, ampliação, rotação e transformação. Essas operações podem ser muito úteis para gerar planos e explicações em domínios nos quais as representações pictóricas se aplicam. O esquema explicativo para a representação visual é:

*Objetivo da explicação:*

- Por que as pessoas têm um tipo particular de comportamento inteligente?

*Padrão explicativo:*

- As pessoas têm imagens visuais de situações.
- As pessoas têm processos tais como escaneamento e rotação que operam sobre essas imagens.
- Os processos para construir e manipular imagens produzem o comportamento inteligente.

As imagens podem ajudar na aprendizagem, e alguns aspectos metafóricos da linguagem podem ter suas raízes em imagens. Alguns experimentos psicológicos sugerem que procedimentos visuais tais como o escaneamento e a rotação empregam imagens, e resultados neurofisiológicos confirmam uma ligação física próxima do raciocínio com imagens mentais e com percepção. Imagens não são apenas visuais, mas podem também operar com outras experiências sensoriais tais como a audição, o tato, o olfato, o paladar, a dor, o equilíbrio, a náusea, a saciação e a emoção.

## 4.6 Conexionismo

Redes conexionistas, que consistem em pontos de conexão e ligações, são muito úteis para entender processos psicológicos que envolvem restrições de satisfação com respeito ao paralelismo. Esses processos incluem aspectos da visão, tomada de decisão, seleção da explicação e produção de sentido na compreensão linguística. Os modelos conexionistas podem simular a aprendizagem por métodos que incluem a aprendizagem hebbiana<sup>3</sup> e a retropropagação. O esquema explicativo para a abordagem conexionista é:

*Objetivo da explicação:*

- Por que as pessoas têm um tipo particular de comportamento inteligente?

*Padrão explicativo:*

- As pessoas têm representações que contêm unidades de processamento simples ligadas umas às outras por conexões excitatórias e inibitórias.
- As pessoas detêm processos que difundem a ativação entre as unidades por meio de suas conexões, bem como processos para modificar essas conexões.
- A aplicação da difusão de ativação e de aprendizagem às unidades produz o comportamento.

Simulações de vários experimentos psicológicos mostraram a relevância psicológica dos modelos conexionistas, que são, contudo, apenas aproximações grosseiras de redes neuronais reais.

## 4.7 Neurociência Teórica

A neurociência teórica é uma tentativa de desenvolver teorias e modelos matemáticos e computacionais de estruturas e processos dos cérebros de humanos

---

<sup>3</sup>N.T.: O psicólogo canadense Donald Hebb (1949) sugeriu que um dos mecanismos de aprendizagem ocorre por reforço sináptico. Quando dois neurônios conectados estão ativados, a conexão se reforça e a probabilidade de que fiquem ativados juntos novamente aumenta.

e de outros animais. Ela difere do conexionismo por tentar ser mais acurada biologicamente, modelando o comportamento de um grande número de neurônios reais organizados em áreas funcionalmente significativas do cérebro. Modelos computacionais do cérebro tornaram-se biologicamente mais ricos, tanto por empregarem neurônios mais realistas que disparam e têm vias químicas quanto por simularem interações entre diferentes áreas do cérebro como o hipocampo e o córtex. Esses modelos não são estritamente uma alternativa às explicações computacionais em termos de lógica, regras, conceitos, analogias, imagens e conexões, mas deveriam se combinar a elas e mostrar como o funcionamento mental pode ser realizado no nível neuronal. O esquema explicativo para a neurociência teórica é:

*Objetivo da explicação:*

- Como o cérebro executa funções tais como as tarefas cognitivas?

*Padrão explicativo:*

- O cérebro tem neurônios organizados em populações e áreas cerebrais por meio de conexões sinápticas.
- As populações neuronais têm padrões de disparos que são transformados por entradas sensoriais e por padrões de disparos de outras populações neuronais.
- As interações das populações neuronais executam funções incluindo as tarefas cognitivas.

Desde a perspectiva da neurociência teórica, as representações mentais são padrões de atividade neuronal, e a inferência é a transformação desses padrões.

## 4.8 Bayesianismo

Os modelos bayesianos são proeminentes na ciência cognitiva, com aplicações a fenômenos psicológicos como a aprendizagem, a visão, o controle motor, a linguagem e a cognição social. Eles também têm aplicações efetivas na robótica. A abordagem bayesiana supõe que a cognição é aproximadamente ideal de acordo com a teoria da probabilidade, especialmente o teorema de Bayes, que afirma que a probabilidade de uma hipótese, dada a evidência, é igual ao resultado de multiplicar a probabilidade anterior da hipótese pela probabilidade condicional da evidência dada a hipótese, tudo isso dividido pela probabilidade da evidência.

O esquema explicativo para a cognição bayesiana é:

*Objetivo da explicação:*

- Como a mente executa funções tais como a inferência?

*Padrão explicativo:*

- A mente tem representações para correlações estatísticas e probabilidades condicionais.
- A mente tem a capacidade para computações probabilísticas tais como as aplicações do teorema de Bayes.
- A aplicação de computações probabilísticas a representações estatísticas executa tarefas mentais tal como a inferência.

Embora os métodos bayesianos tenham tido aplicações impressionantes em uma vasta gama de fenômenos, a sua plausibilidade psicológica é discutível por causa das suposições sobre otimização e computação baseadas na teoria da probabilidade.

## 4.9 Aprendizagem Profunda

A inteligência artificial tem sido uma parte central da ciência cognitiva desde os anos 50. Os mais dramáticos avanços recentes na IA vieram da abordagem da aprendizagem profunda, a qual produziu importantes descobertas em áreas que incluem jogos, reconhecimento de objetos e tradução. A aprendizagem profunda se apoia em ideias do conexionismo e da neurociência teórica, mas usa redes neuronais com mais camadas e algoritmos aprimorados, beneficiando-se de computadores mais velozes e grandes bases de dados de exemplos. Outra inovação importante é a combinação da aprendizagem, por exemplo, e a aprendizagem por reforço, o que resultou, em 2016, no jogador líder mundial de Go, AlphaGo. As ideias da aprendizagem profunda estão reverberando de volta na neurociência e também começam a influenciar a pesquisa na psicologia cognitiva. O esquema explicativo para a aprendizagem profunda é:

*Objetivo da explicação:*

- Como o cérebro executa funções tais como as tarefas cognitivas?



*Padrão explicativo:*

- O cérebro tem um número grande de neurônios organizados em 6-20 camadas.
- O cérebro tem mecanismos poderosos para aprendizagem por exemplos e para aprender ações que são reforçadas pelos seus sucessos.
- A aplicação de mecanismos de aprendizagem a redes neuronais em camadas as torna capazes de desempenho humano e algumas vezes mesmo sobre-humano.

Embora a aprendizagem profunda tenha produzido melhoras dramáticas em alguns sistemas de IA, não é claro como ela pode ser aplicada a aspectos do pensamento humano que inclui imagens, emoções e analogia.

## 5. Relevância Filosófica

Alguma filosofia, em particular a filosofia naturalista da mente, faz parte da ciência cognitiva. Porém o campo interdisciplinar da ciência cognitiva é relevante para a filosofia de diferentes maneiras. Primeiro, resultados psicológicos, computacionais e outros das investigações na ciência cognitiva têm importantes aplicações possíveis em problemas filosóficos tradicionais da epistemologia, da metafísica e da ética. Segundo, a ciência cognitiva pode servir como um objeto de crítica filosófica, particularmente no que diz respeito à suposição central de que o pensamento é representacional e computacional. Terceiro e mais construtivamente, a ciência cognitiva pode ser considerada como um objeto de investigação na filosofia da ciência, gerando reflexões sobre a metodologia e as pressuposições do empreendimento.

### 5.1 Aplicações Filosóficas

Boa parte da pesquisa filosófica hoje em dia é naturalista, ao tratar as investigações filosóficas como contínuas ao trabalho empírico em áreas tais como a psicologia. De uma perspectiva naturalista, a filosofia da mente é aliada próxima do trabalho teórico e experimental da ciência cognitiva. As conclusões metafísicas sobre a natureza da mente devem ser obtidas não por especulação *a priori*, mas

por reflexão informada pelos desenvolvimentos científicos em campos como a psicologia, a neurociência e a ciência da computação. De modo similar, a epistemologia não é um exercício conceitual autônomo, ela depende e se beneficia de descobertas científicas que dizem respeito a estruturas mentais e a processos de aprendizagem. A ética pode se beneficiar ao usar um entendimento ampliado da psicologia do pensamento moral para tratar de questões éticas tais como a natureza de deliberações sobre o certo e o errado. Abaixo elenco alguns problemas filosóficos para os quais os desenvolvimentos em curso na ciência cognitiva são muito relevantes. Oferecemos ligações para outros artigos relevantes nesta Enciclopédia:

- *Inatismo*. Até que ponto o conhecimento é inato ou adquirido pela experiência? O comportamento humano é moldado primariamente pela natureza ou pela educação?
- *Linguagem do pensamento*. O cérebro humano opera com um código parecido com a linguagem ou com uma arquitetura conexionista mais geral? Qual é a relação entre modelos cognitivos simbólicos que utilizam regras e conceitos e modelos sub-simbólicos que utilizam redes neuronais?
- *Imagens mentais*. As mentes humanas pensam com imagens visuais ou de outros tipos, ou apenas com representações parecidas com a linguagem?
- *Psicologia popular*. A compreensão ordinária de uma pessoa acerca de outras pessoas consiste em ter uma teoria da mente ou apenas em ser capaz de simulá-las?
- *Significado*. Como representações mentais adquirem significado ou conteúdo mental? Até que ponto o significado de uma representação depende da sua relação com outras representações, da sua relação com o mundo e da sua relação com a comunidade de pensantes?
- *Identidade mente-cérebro*. Os estados mentais são estados cerebrais? Eles podem ser multiplamente realizados por outros estados materiais? Qual é a relação entre a psicologia e a neurociência? O materialismo é verdadeiro?

- *Livre-arbítrio*. A ação humana é livre ou meramente causada por eventos cerebrais?
  - *Psicologia moral*. Como mentes/cérebros realizam juízos éticos?
  - *O significado da vida*. Como as mentes interpretadas naturalisticamente como cérebros encontram valor e significado?
  - *Emoções*. O que são as emoções, e que papel elas desempenham no pensamento?
  - *Doença mental*. O que são as doenças mentais e como são os processos psicológicos e neuronais relevantes para a sua explicação e seu tratamento?
  - *Percepção e realidade*. Como mentes/cérebros formam e avaliam representações do mundo exterior?
  - *Ciência social*. Como as explicações das operações das mentes interagem com explicações das operações de grupos e sociedades?
- Problemas filosóficos adicionais surgem ao examinar as pressuposições das abordagens atuais da ciência cognitiva.

## 5.2 Crítica à Ciência Cognitiva

A afirmação de que as mentes humanas funcionam por representação e computação é uma conjectura empírica que pode estar errada. Embora a abordagem computacional-representacional da ciência cognitiva tenha sido bem-sucedida na explicação de muitos aspectos da solução de problemas, da aprendizagem e do uso da linguagem humanos, algumas críticas filosóficas têm afirmado que esta abordagem está fundamentalmente errada. Os críticos da ciência cognitiva apresentaram objeções tais como:

- 1 A objeção da emoção: a ciência cognitiva negligencia o papel importante das emoções no pensamento humano.
- 2 A objeção da consciência: a ciência cognitiva ignora a importância da consciência no pensamento humano.

- 3 A objeção do mundo: a ciência cognitiva desconsidera o papel significativo dos ambientes físicos no pensamento humano, o qual é integrado e se estende no mundo.
- 4 A objeção do corpo: a ciência cognitiva negligencia a contribuição da corporificação no pensamento humano e na ação.
- 5 A objeção dos sistemas dinâmicos: a mente é um sistema dinâmico, não um sistema computacional.
- 6 A objeção do social: o pensamento humano é inerentemente social de formas ignoradas pela ciência cognitiva.
- 7 A objeção da matemática: os resultados matemáticos mostram que o pensamento humano não pode ser computacional no sentido padrão. Portanto, o cérebro deve funcionar de modo diferente, talvez como um computador quântico.

As primeiras cinco objeções estão sendo cada vez mais enfrentadas por avanços que explicam as emoções, a consciência, a ação e a corporificação em termos de mecanismos neuronais. A objeção do social está sendo respondida pelo desenvolvimento de modelos computacionais de agentes em interação. A objeção da matemática está baseada em uma má compreensão do teorema de Gödel e em uma ênfase exagerada da relevância da teoria quântica para processos neuronais.

### **5.3 A Filosofia da Ciência Cognitiva**

A ciência cognitiva levanta muitas questões metodológicas interessantes que merecem ser investigadas por filósofos da ciência. Qual é a natureza da representação? Que papel os modelos computacionais desempenham no desenvolvimento de teorias cognitivas? Qual é a relação entre explicações da mente aparentemente concorrentes envolvendo processamento simbólico, redes neuronais e sistemas dinâmicos? Qual é a relação entre os vários campos da ciência cognitiva tais como a psicologia, a linguística e a neurociência? Os fenômenos psicológicos estão sujeitos a explicações reducionistas por meio da neurociência? Os níveis de explicação são melhor caracterizados em termos de níveis ontológicos (molecular, neuronal, psicológico,

social) ou de níveis metodológicos (computacional, algoritmo, físico)?

A proeminência crescente das explicações neuronais na psicologia cognitiva, social, do desenvolvimento e clínica levanta questões filosóficas importantes sobre a explicação e a redução. O antirreduccionismo, de acordo com o qual as explicações psicológicas são completamente independentes das neurológicas, está se tornando cada vez mais implausível, mas permanece controverso até que ponto a psicologia pode ser reduzida à neurociência e à biologia molecular. Para responder questões sobre a natureza da redução é crucial ter respostas para questões sobre a natureza da explicação. As explicações na psicologia, na neurociência e na biologia são plausivelmente, em geral, vistas como descrições de **mecanismos**, que são combinações de partes conectadas que interagem para produzir mudanças regulares. Nas explicações psicológicas, as partes são representações mentais que interagem por meio de procedimentos computacionais para produzir novas representações. Nas explicações neurocientíficas, as partes são populações neuronais que interagem por processos eletroquímicos para produzir nova atividade neuronal que leva a ações. Se o progresso na neurociência teórica continuar, deverá tornar-se possível vincular explicações psicológicas às neurológicas mostrando como as representações mentais tais como conceitos são constituídas por atividades nas populações neuronais, e como os procedimentos computacionais tal como a difusão de ativação entre conceitos pode ser realizada por processos neuronais.

A integração crescente da psicologia cognitiva com a neurociência fornece evidência para a teoria da identidade mente-cérebro de acordo com a qual os processos mentais são neuronais, representacionais e computacionais. Outros filósofos disputam essa identificação com base em que mentes são corporificadas em sistemas biológicos e se estendem no mundo. Contudo, afirmações moderadas sobre a corporificação são consistentes com a teoria da identidade porque as representações cerebrais operam em várias modalidades (e.g. visual e motora) que permitem às mentes lidar com o mundo. Outra alternativa materialista à identidade mente-cérebro vem do reconhecimento de que as explicações da mente empregam mecanismos moleculares e sociais tanto quanto neuronais e representacionais.

## Referência bibliográfica

ANDERSON, J. **How Can the Mind Occur in the Physical Universe?** Oxford: Oxford University Press, 2007.

- ANDERSON, J. **Cognitive Psychology and its Implications**. New York: Worth, 2010.
- BECHTEL, W. **Mental Mechanisms: Philosophical Perspectives on Cognitive Neurosciences**. New York: Routledge, 2008.
- BECHTEL, W.; GRAHAM, G. (ed.). **A Companion to Cognitive Science**. Malden, MA: Blackwell, 1998.
- BECHTEL, W. *et al.* (ed.). **Philosophy and the Neurosciences: A Reader**. Malden, MA: Blackwell, 2001.
- BERMÚDEZ, J. **Cognitive Science: An Introduction to the Science of the Mind**. 2nd. ed. Cambridge: Cambridge University Press, 2014.
- BLOUW, P. *et al.* Concepts as Semantic Pointers: A Framework and Computational Model. **Cognitive Science**, v. 40, p. 1128-1162, 2016.
- BODEN, M. **Mind as Machine: A History of Cognitive Science**. Oxford: Clarendon, 2006.
- CHEMERO, A. **Radical Embodied Cognitive Science**. Cambridge, MA: MIT Press, 2009.
- CHURCHLAND, P. **Brain-wise: Studies in Neurophilosophy**. Cambridge, MA: MIT Press, 2002.
- CHURCHLAND, P. **Neurophilosophy at Work**. Cambridge: Cambridge University Press, 2007.
- CLARK, A. **Mindware: An Introduction to the Philosophy of Cognitive science**. New York: Oxford University Press, 2001.
- CLARK, A. **Supersizing the Mind: Embodiment, Action, and Cognitive Extension**. New York: Oxford University Press, 2008.
- CRAVER, C. **Explaining the Brain**. Oxford: Oxford University Press, 2007.
- DAYAN, P.; ABBOTT, L. **Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems**. Cambridge, MA: MIT Press, 2001.
- DEHAENE, S. **Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts**. New York: Viking, 2014.
- DREYFUS, H. **What Computers Still Can't Do**. Cambridge, MA: MIT Press, 1992.
- ELIASMITH, C. **How to Build a Brain: A Neural Architecture for Biological Cognition**. Oxford: Oxford University Press, 2013.
- ELIASMITH, C.; ANDERSON, C. **Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems**. Cambridge, MA: MIT Press, 2003.
- FORBUS, K. *et al.* Extending SME to Handle Larger-Scale Cognitive modeling. **Cognitive Science**, v. 41, n. 5, p. 1152-1201, 2017.
- FRIEDENBERG, J.; SILVERMAN, G. **Cognitive Science: An Introduction to the Study of Mind**. Thousand Oaks, CA: Sage, 2011.

- GIBBS, R. **Embodiment and Cognitive Science**. Cambridge: Cambridge University Press, 2005.
- GOLDMAN, A. **Philosophical Applications of Cognitive Science**. Boulder: Westview Press, 1993.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, **Deep Learning**. Cambridge, MA: MIT Press, 2016.
- GRIFFITHS, T.; KEMP, C.; TENENBAUM, J. Bayesian Models of Cognition. In: SUN, R. (ed.). **The Cambridge Handbook of Computational Psychology**. Cambridge: Cambridge University Press, 2008. p. 59-100.
- HEBB, D. **The Organization of Behavior: a Neuropsychological Theory**. London: Lawrence Erlbaum Associates, Publishers, 1949.
- HOFSTADTER, D.; SANDER, E. **Surfaces and Essences: Analogy as the Fuel and Fire of Thinking**. New York: Basic Books, 2013.
- HOLYOAK, K.; MORRISON, R. (ed.). **The Oxford Handbook of Thinking and Reasoning**. New York: Oxford University Press, 2012.
- KNOBE, J.; NICHOLS, S. (ed.). **Experimental Philosophy**. Oxford: Oxford University Press, 2008.
- KOSSLYN, S.; THOMPSON, W.; GANIS, G. **The Case for Mental Imagery**. New York: Oxford University Press, 2016.
- LAIRD, J.; LEBIERE, C.; ROSENBLOOM, P. A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. **AI Magazine**, v. 38, n. 4, p. 13-26, 2017.
- LECUN, Y.; BENGIO, Y.; HINTON, Y. Deep Learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015.
- MARGOLIS, E.; LAURENCE, S. (ed.). **The Conceptual Mind: New Directions in the Study of Concepts**. Cambridge, MA: MIT Press, 2015.
- MCCAULEY, R. Reduction: Models of Cross-scientific Relations and their Implications for the Psychology-neuroscience Interface. In: THAGARD, P. (Ed.). **Philosophy of Psychology and Cognitive Science**. Amsterdam: Elsevier, 2007. p. 105-158.
- MILKOWSKI, M. **Explaining the Computational Mind**. Cambridge, MA: MIT Press, 2013.
- MURPHY, D. **Psychiatry in the Scientific Image**. Cambridge, MA: MIT Press, 2006.
- NADEL, L. (ed.). **Encyclopedia of Cognitive Science**. London: Nature Publishing Group, 2003.
- NISBETT, R. **The Geography of Thought: How Asians and Westerners Think Differently ... and Why**. New York: Free Press, 2003.

- O'REILLY, R. *et al.*, 2012. Computational Cognitive Neuroscience, Wiki Book. Disponível em: <http://ccnbook.colorado.edu>. Acesso em: 10 jan. 2022.
- PEARL, J.; MACKENZIE, D. **The Book of Why: The New Science of Cause and Effect**. New York: Basic Books, 2018.
- PESSOA, L. **The Cognitive-Emotional Brain: From Interactions to Integration**. Cambridge, MA: MIT Press, 2013.
- PIANTADOSI, S.; TENENBAUM, J.; GOODMAN, N. The Logical Primitives of Thought: Empirical Foundations for Compositional Cognitive Models. **Psychological Review**, v. 123, n. 4, p. 392-424, 2015.
- POLK, T.; SEIFERT, C. (ed.). **Cognitive Modeling**. Cambridge, MA: MIT Press, 2002.
- ROGERS, T.; MCCLELLAND, J. **Semantic Cognition: A Parallel Distributed Processing Approach**. Cambridge, MA: MIT Press, 2004.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Upper Saddle River, NJ: Prentice Hall, 2009.
- SMITH, E.; KOSSLYN, S. **Cognitive Psychology: Mind and Brain**. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.
- SOBEL, C. **The Cognitive Sciences: An Interdisciplinary Approach**. Mountain View, CA: Mayfield, 2001.
- SUN, R. (ed.). **The Cambridge Handbook of Computational Psychology**. Cambridge: Cambridge University Press, 2008.
- SUN, R. (ed.). **Grounding Social Sciences in Cognitive Sciences**. Cambridge, MA: MIT Press, 2012a.
- THAGARD, P. **Mind: Introduction to Cognitive Science**. second ed. Cambridge, MA: MIT Press, 2005.
- THAGARD, P. (ed.). **Philosophy of Psychology and Cognitive Science**. Amsterdam: Elsevier, 2007.
- THAGARD, P. Why cognitive science needs philosophy and vice versa. **Topics in Cognitive Science**, v. 1, p. 237-254, 2009.
- THAGARD, P. **The Brain and the Meaning of Life**. Princeton: Princeton University Press, 2010.
- THAGARD, P. **The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change**. Cambridge, MA: MIT Press, 2012.
- THAGARD, P. **Brain-Mind: From Neurons to Consciousness and Creativity**. Oxford: Oxford University Press, 2019.
- THOMPSON, E. **Mind in Life: Biology, Phenomenology, and the Science of Mind**. Cambridge, MA: Harvard University Press, 2007.



VON ECKARDT, B. **What is Cognitive Science?** Cambridge, MA: MIT Press, 1993.  
WILSON, R.; KEIL, F. (ed.). **The MIT Encyclopedia of the Cognitive Sciences.**  
Cambridge, MA: MIT Press, 1999.

## (II) A Teoria Computacional da Mente\*

Autor: Michael Rescorla

Tradução: Jessica Nunes Vergara (IFSUL/PPGFIL-UFRGS)

Revisão: César Fernando Meurer (UFABC)

Uma máquina poderia pensar? Poderia a própria mente ser uma máquina pensante? A revolução computacional transformou a discussão dessas questões, oferecendo nossas melhores candidatas até agora para máquinas que emulam raciocínio, tomada de decisão, resolução de problemas, percepção, compreensão linguística e outros processos mentais. Avanços em computação levantam a hipótese de que a mente é um sistema computacional, uma concepção conhecida como a *teoria computacional da mente* (TCM). **Computacionalistas** são pesquisadores que endossam a TCM, ao menos quanto a certos processos mentais importantes. A TCM desempenhou um papel central nas ciências cognitivas durante as décadas de 1960 e 1970. Por muitos anos, desfrutou de caráter ortodoxo. Mais recentemente,

---

\*RESCORLA, M. The Computational Theory of Mind. In: ZALTA, E. N. (ed.). Stanford Encyclopedia of Philosophy. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: <https://plato.stanford.edu/archives/spr2020/entries/computational-mind/>. Acesso em: 10 out. 2021.

The following is the translation of the entry on The Computational Theory of Mind by Michael Rescorla, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/spr2020/entries/computational-mind/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/computational-mind/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.

tem sido pressionada por vários paradigmas rivais. Uma tarefa essencial enfrentada pelos computacionalistas é a de explicar o que se pretende dizer quando se diz que a mente “computa”. Uma segunda tarefa é a de defender que a mente “computa” em um sentido relevante do termo. Uma terceira tarefa é a de elucidar como as descrições computacionais se relacionam com outros tipos comuns de descrições, especialmente *descrições neurofisiológicas* (as quais utilizam propriedades neurofisiológicas do corpo ou do cérebro do organismo) e *descrições intencionais* (as quais utilizam propriedades representacionais dos estados mentais).

## 1. Máquinas de Turing

As noções intuitivas de **computação e algoritmo** são centrais à matemática. Grosso modo, um algoritmo é um procedimento passo a passo explícito para responder alguma questão ou solucionar algum problema. Um algoritmo provê *instruções mecânicas de rotina* prescrevendo como proceder a cada passo. Obedecer às instruções não requer qualquer habilidade especial ou criatividade. Por exemplo, os algoritmos familiares ensinados nos primeiros anos do ensino fundamental descrevem como computar a adição, a multiplicação e a divisão. Até o início do século XX, os matemáticos se apoiavam em noções informais de computação e algoritmo sem nenhuma tentativa de realizar qualquer coisa próxima a uma análise formal. Progressos nas fundamentações da matemática eventualmente impeliram lógicos a buscar um tratamento mais sistemático. O artigo divisor de águas de Alan Turing (1936), *On Computable Numbers, With an Application to the Entscheidungsproblem*<sup>4</sup> ofereceu a análise que se mostrou mais influente.

Uma *máquina de Turing* é um modelo abstrato de um dispositivo computacional ideal com tempo e espaço de armazenamento ilimitados à sua disposição. O dispositivo manipula **símbolos**, da mesma forma que um agente humano que computa manipulando riscos de lápis num papel durante uma computação aritmética. Turing diz muito pouco sobre a natureza dos símbolos. Ele assume que símbolos primitivos são retirados de um alfabeto finito. Ele também assume que símbolos podem ser inscritos e apagados nos “locais de memória”. O modelo de Turing trabalha da seguinte forma:

---

<sup>4</sup> N.T.: “Dos números computáveis, com uma aplicação ao Problema da Decisão”.

- Há locais de memória infinitos, arranjados numa estrutura linear. Metaforicamente, esses locais de memória são “células” numa “fita de papel” infinitamente longa. Colocando de uma forma mais literal, os locais de memória podem ser realizados fisicamente em diferentes mídias (por exemplo, *chips* de silício).
- Há um processador central, o qual pode acessar um local de memória por vez. Metaforicamente, o processador central é um escâner (*scanner*) que se move ao longo da fita de papel de “célula” em “célula”, uma por vez.
- O processador central pode entrar em finitamente muitos **estados de máquina** (*machine states*).
- O processador central pode executar quatro operações elementares: escrever um símbolo num local de memória; apagar um símbolo de um local de memória; acessar o próximo local de memória no arranjo linear (“mover-se para a direita da fita”); acessar o local de memória anterior no arranjo linear (“mover-se para a esquerda na fita”).
- Qual operação elementar o processador central executa depende inteiramente de dois fatos: qual símbolo está atualmente inscrito no local de memória atual e o estado de máquina atual do próprio escâner.
- Uma **tabela de máquina** dita qual operação elementar o processador executa, dado o seu estado de máquina atual e o símbolo que está atualmente acessando. A tabela de máquina também dita como o estado de máquina do processador central muda dados estes mesmos fatores. Desse modo, a tabela de máquina engloba um conjunto finito de instruções mecânicas de rotina governando a computação.

Turing traduz essa descrição informal para um modelo matemático rigoroso. Para mais detalhes, consulte o verbete da SEP **Turing machine**.

Turing inspira sua abordagem na reflexão sobre agentes humanos ideais. Mencionando limites finitários de nosso aparato perceptual e cognitivo, ele argumenta

que qualquer algoritmo simbólico executado por um humano pode ser replicado por uma máquina de Turing adequada. Ele conclui que o formalismo da máquina de Turing, apesar de sua simplicidade extrema, é poderoso o suficiente para capturar todos os procedimentos mecânicos sobre configurações simbólicas humanamente executáveis. Quase a totalidade de seus interlocutores subsequentes concordaram com ele.

A computação de Turing é frequentemente descrita como **digital** em oposição à **analógica**. O que isso significa não é sempre tão claro, mas a ideia básica da descrição digital é normalmente a de que a computação opera sobre configurações discretas. Comparativamente, muitos algoritmos historicamente importantes operam sobre configurações que variam de forma contínua. Por exemplo, a geometria euclidiana dá um papel enorme para *construções de régua e compasso*, as quais manipulam formas geométricas. Para cada forma, pode-se encontrar outra que difere dela em uma pequena extensão arbitrariamente estipulada. Configurações simbólicas manipuladas por uma máquina de Turing não se diferem por pequenas extensões estipuladas arbitrariamente. Máquinas de Turing operam sobre cadeias de elementos discretos (dígitos) retirados de um alfabeto finito. Uma controvérsia recorrente diz respeito a se o paradigma digital é satisfatoriamente adequado para modelar a atividade mental ou se um paradigma analógico seria mais adequado (MACLENNAN, 2012; PICCINI; BAHAR, 2013)

Além de introduzir as máquinas de Turing, Turing (1936) provou inúmeros resultados matemáticos seminais as envolvendo. Em particular, ele provou a existência de uma *máquina de Turing universal* (MTU). Grosso modo, uma MTU é uma máquina de Turing que pode mimetizar qualquer outra máquina de Turing. Mune-se a MTU com uma entrada simbólica que codifica a tabela da máquina de Turing *M*. A MTU replica o comportamento de *M*, executando instruções prescritas pela tabela de máquina de *M*. Nesse sentido, a MTU é um *computador programável de uso geral*. Numa primeira aproximação, todos os computadores pessoais são de uso geral: eles podem mimetizar qualquer máquina de Turing, quando adequadamente programados. A principal ressalva é a de que computadores físicos têm memória finita, enquanto a máquina de Turing tem memória ilimitada. Mais precisamente, então, um computador pessoal pode mimetizar qualquer máquina de Turing **até que seu suprimento limitado de memória seja exaurido**.

A discussão de Turing ajudou a estabelecer as bases da *ciência da computação*, a qual objetiva desenhar, construir e compreender sistemas computacionais. Como nós sabemos, cientistas da computação podem hoje construir máquinas computacionais extremamente sofisticadas. Todas essas máquinas implementam

algo parecido com a computação de Turing, embora nos detalhes se diferenciem do modelo simplificado de Turing.

## 2. Inteligência artificial

O rápido progresso na ciência da computação incitou muitos, incluindo Turing, a se perguntar se poderíamos construir um computador capaz de pensar. A *Inteligência Artificial* (IA) visa construir “maquinarias pensantes”. Mais precisamente, visa construir máquinas computacionais que executam tarefas mentais essenciais, tais como raciocínio, tomada de decisão, resolução de problemas e assim por diante. Durante as décadas de 1950 e 1960, esse objetivo passou a parecer cada vez mais realista (HAUGELAND, 1985).

Estudos iniciais em IA enfatizavam a **lógica**. Os pesquisadores visavam “mecanizar” o raciocínio dedutivo. Um exemplo famoso foi o programa de computador *Logic Theorist* (NEWELL; SIMON, 1956, o qual provou 38 dos primeiros 52 teoremas do *Principia Mathematica* (WHITEHEAD; RUSSELL, 1925). Em um dos casos, encontrou uma prova mais simples que a do *Principia*.

Sucesso inicial desse tipo estimulou um interesse imenso dentro e fora da academia. Muitos pesquisadores previram que em poucos anos existiriam máquinas inteligentes. Obviamente, essas previsões não se realizaram. Robôs inteligentes ainda não caminham entre nós. Mesmo processos mentais de nível relativamente baixo, tais como a percepção, excedem de longe as capacidades dos atuais programas de computador. Quando as previsões confiantes acerca de máquinas pensantes se provaram otimistas demais, muitos observadores perderam o interesse ou concluíram que a IA era uma missão tola. No entanto, as décadas têm testemunhado um progresso gradual. Um sucesso impressionante foi o *Deep Blue*, da IBM, que derrotou o campeão de xadrez Gary Kasparov em 1997. Outro grande sucesso foi o carro sem motorista, Stanley (THRUN; MONTEMERLO; DAHLKAMP *et al.*, 2006), que completou um percurso de 132 milhas no Deserto de Mojave, vencendo o Grande Desafio da *Defense Advanced Research Projects Agency* (DARPA)<sup>5</sup> em 2005. Uma história de sucesso com menos holofotes é a grande melhoria nos

---

<sup>5</sup> N.T.: Agência de Projetos de Pesquisa Avançada de Defesa.

algoritmos de reconhecimento de voz.

Um problema que atrapalhou o trabalho inicial em IA foi a *incerteza*. Quase todo raciocínio e tomada de decisão opera sob condições de incerteza. Por exemplo, você pode precisar decidir se irá a um piquenique mesmo estando incerto acerca de se irá chover. A *teoria bayesiana da decisão* é o modelo matemático padrão para inferência e tomada de decisão realizados sob incerteza. A incerteza é codificada em termos de *probabilidade*. Regras precisas ditam como atualizar as probabilidades à luz de novas evidências e como selecionar ações à luz das probabilidades e utilidades. (Veja os verbetes da SEP “Bayes’s theorem” e “normative theories of rational choice: expected utility” para mais detalhes.) Nas décadas de 1980 e 1990, desenvolvimentos tecnológicos e conceituais possibilitaram programas de computador eficientes que implementaram, ou se aproximaram de implementar, a inferência bayesiana em cenários realistas. Seguiu-se uma explosão de IA Bayesiana (THRUN; BURGARD; FOX, 2006), incluindo os avanços anteriormente mencionados acerca do reconhecimento de fala e de veículos sem motorista. Algoritmos tratáveis que lidam com a incerteza são uma grande conquista da IA contemporânea (MURPHY, 2012) e, possivelmente, um prenúncio de um progresso futuro ainda mais impressionante.

Alguns filósofos insistem que computadores, não importa o quão sofisticados se tornem, irão, na melhor das hipóteses, **mimetizar** o pensamento, mas não o **replicar**. Em uma simulação meteorológica feita em computador não chove realmente. Uma simulação de um voo feita em computador não voa realmente. Mesmo que um sistema computacional pudesse simular uma atividade mental, por que pensar que seria um caso genuíno de um evento desse tipo?

Turing (1950) antecipou essas preocupações e tentou neutralizá-las. Ele propôs um cenário, agora conhecido como o *Teste de Turing*, no qual se avalia se um interlocutor oculto é um computador ou um humano. Um computador *passa no teste de Turing* se não se puder estabelecer que ele é um computador. Turing propôs que abandonássemos a pergunta “um computador pode pensar?” uma vez que seria desanimadoramente vaga, substituindo-a por “um computador poderia passar no teste de Turing?”. A discussão proposta por Turing recebeu uma atenção considerável, provando-se especialmente influente dentro do campo de IA. Ned Block (1981) ofereceu uma crítica influente. Ele argumentou que algumas máquinas possíveis passariam no teste de Turing mesmo que essas máquinas não chegassem perto de possuir pensamentos ou inteligência genuínos. Veja o verbe da SEP, para uma discussão da objeção de Block e outros problemas que circundam o Teste de Turing.

Mais sobre IA pode ser encontrado no verbete , na SEP. Para mais detalhes, confira Russell e Norvig (2010).

### 3. A teoria clássica computacional da mente

Warren McCulloch e Walter Pitts (1943) foram os primeiros a sugerir que algo que parecesse uma máquina de Turing talvez fosse um bom modelo para a mente. Na década de 1960, a computação de Turing tornou-se central para a iniciativa interdisciplinar que estava emergindo, as *ciências cognitivas*, a qual estuda a mente a partir da psicologia, da ciência da computação (especialmente IA), da linguística, da filosofia, da economia (especialmente da teoria dos jogos (*game theory*) e da economia comportamental (*behavioral economics*)), da antropologia e da neurociência. O nome “teoria clássica computacional da mente” (a qual abreviaremos como TCCM) é hoje relativamente padrão. De acordo com a TCCM, a mente é um sistema computacional similar em pontos importantes a uma máquina de Turing e processos mentais centrais (por exemplo, raciocínio, tomada de decisão e resolução de problemas) são computações similares em pontos importantes a computações executadas por uma máquina de Turing. Essas formulações são imprecisas. A TCCM é melhor compreendida como uma família de concepções, ao invés de uma concepção unitária e bem definida<sup>6</sup>.

É comum descrever a TCCM como incorporando a “metáfora do computador”. Essa descrição é duplamente enganosa.

Primeiramente, a TCCM é melhor formulada como descrevendo a mente como um “sistema que computa” ou como um “sistema computacional” ao invés de como um “computador”. Como David Chalmers (2011) apontou, descrever um sistema como um “computador” sugere fortemente que o sistema é **programável**.

---

<sup>6</sup> O rótulo “clássico” é, por vezes, tido como incluindo doutrinas adicionais, para além da tese central segundo a qual a atividade mental é computação ao estilo Turing: por exemplo, que a computação mental manipula símbolos com conteúdo representacional ou que a computação mental manipula representações mentais com estrutura de constituintes de estilo parte/todo ou, ainda, que a computação mental instancia algo como a arquitetura de Von Neumann para computadores digitais. Note, também, que a abreviação “TCCM” é, às vezes, utilizada como sigla para a *teoria computacional conexionista da mente*.



Como Chalmers também apontou, não se está obrigado a alegar que a mente é programável simplesmente porque se entende a mente como um sistema computacional ao estilo Turing. (A maioria das máquinas de Turing não são programáveis). Portanto, a expressão “metáfora do computador” fortemente sugere um compromisso teórico que não é essencial à TCCM. O ponto aqui não é meramente terminológico. Críticos da TCCM geralmente objetam argumentando que a mente não é um computador programável de uso geral (CHURCHLAND; KOCH; SEJNOWSKI, 1990). Uma vez que os computacionalistas clássicos não precisam alegar (e normalmente não alegam) que a mente é um computador programável de uso geral, a objeção é mal direcionada.

Em segundo lugar, a TCCM não foi pensada para ser metafórica. A TCCM não defende simplesmente que a mente é **como** um sistema que computa. A TCCM defende que a mente é **literalmente** um sistema que computa. Obviamente, os sistemas artificiais de computação mais familiares são feitos de *chips* de silício ou materiais semelhantes, enquanto o corpo humano é feito de carne e sangue. Mas a TCCM defende que essa diferença esconde uma similaridade mais fundamental, a qual podemos capturar por meio de um modelo computacional ao estilo Turing. Ao oferecer tal modelo, nós prescindimos dos detalhes físicos. Nós alcançamos uma descrição computacional abstrata que poderia ser fisicamente implementada em diversos meios (por exemplo, por meio de *chips* de silício, ou de neurônios, ou de roldanas e alavancas). A TCCM defende que um modelo computacional abstrato adequado oferece uma descrição literalmente correta dos processos mentais essenciais.

É comum resumir a TCCM com o *slogan* “a mente é uma máquina de Turing”. Esse *slogan* é também um tanto enganosa, pois ninguém considera o formalismo preciso de Turing como um modelo plausível de atividade mental. O formalismo parece restritivo demais por muitos motivos:

- As máquinas de Turing executam computação simbólica pura. As entradas (*inputs*) e as saídas (*outputs*) são símbolos inscritos nos locais de memória. A mente, por contraste, recebe **entradas sensórias** (por exemplo, estímulos retiniais) e produz **saídas motoras** (por exemplo, ativações musculares). Uma teoria completa precisa descrever como a computação mental interage com as entradas sensórias e as saídas motoras.
- Uma máquina de Turing tem capacidade infinita de memória discreta. Sistemas biológicos ordinários têm

capacidade finita de memória. Um modelo psicológico plausível deve substituir a disponibilidade infinita de memória por uma disponibilidade de memória que seja grande, mas finita.

- Computadores modernos têm **memória de acesso aleatório**: locais de memória endereçáveis que o processador central pode acessar diretamente. Memórias de máquinas de Turing não são endereçáveis. O processador central pode acessar um local apenas acessando sequencialmente os locais intermediários. A computação sem memória endereçável é irremediavelmente ineficiente. Por essa razão, C. R. Gallistel e Adam King (2009) argumentam que a memória endereçável fornece um melhor modelo da mente do que a memória não endereçável.
- Uma máquina de Turing tem um processador central que opera de **forma seriada**, executando uma instrução por vez. Outros formalismos computacionais flexibilizam essa presunção, permitindo várias unidades de processamento que operam em **paralelo**. Computacionalistas clássicos podem aceitar computações paralelas (FODOR; PYLYSHYN, 1988; GALLISTEL; KING, 2009, p. 174). Veja Gandy (1980) e Sieg (2009) para tratamentos matemáticos gerais que abrangem tanto a computação serial quanto a paralela.
- A computação de Turing é **determinística**: o estado total computacional determina o estado computacional subsequente. Pode-se, ao invés, permitir computações **estocásticas**. Em um modelo estocástico, o estado atual não determina um estado posterior único. Em vez disso, existe uma certa probabilidade de que a máquina fará a transição de um estado para outro.

A TCCM alega que a atividade mental é “computação ao estilo Turing”, permitindo esses e outros distanciamentos do formalismo do próprio Turing.

### 3.1 Funcionalismo de máquina

Hilary Putnam (1967) introduziu a TCCM na filosofia. Ele contrastou sua posição com o *behaviorismo lógico* e com a *teoria tipo-identidade* (*type-identity theory*). Cada teoria pretende revelar a natureza dos estados mentais, incluindo as atitudes proposicionais (por exemplo, crenças), as sensações (por exemplo, dores) e as emoções (por exemplo, medo). De acordo com o behaviorismo lógico, estados mentais são disposições comportamentais. De acordo com a teoria tipo-identidade, estados mentais são estados cerebrais. Putnam avançou uma visão **funcionalista** alternativa, para a qual estados mentais são estados funcionais. De acordo com o funcionalismo, um sistema tem uma mente quando o sistema tem uma **organização funcional** adequada. Estados mentais são estados que desempenham papéis apropriados na organização funcional do sistema. Cada estado mental é individuado por suas interações com entradas sensoriais, saídas motoras e outros estados mentais.

O funcionalismo oferece vantagens significativas sobre o behaviorismo lógico e a teoria tipo-identidade:

- Behavioristas querem associar cada estado mental com um padrão de comportamento característico - uma tarefa desanimadora, pois estados mentais individuais comumente não têm efeitos comportamentais característicos. O comportamento quase sempre resulta de estados mentais distintos operando conjuntamente (por exemplo, uma crença e um desejo). O funcionalismo evita essa dificuldade ao individuar estados mentais por meio de relações características não apenas entre as entradas sensoriais e o comportamento, mas também entre estados mentais eles mesmos.
- Teóricos de tipo-identidade querem associar cada estado mental a uma característica física ou um estado neurofisiológico. Putnam coloca esse projeto em dúvida ao argumentar que estados mentais são **multiplamente realizáveis**: o mesmo estado mental pode ser realizado por sistemas físicos diversos, incluindo não apenas criaturas terrestres, mas também criaturas hipotéticas (por exemplo, um marciano à base de silício). O funcionalismo é confeccionado

sob medida para acomodar a múltipla realizabilidade (*multiple realizability*). De acordo com o funcionalismo, o que importa para a mentalidade é um padrão de organização, o qual pode ser fisicamente implementado de muitas maneiras distintas. Veja o verbete da SEP, **multiple realizability**, para mais discussões a respeito desse argumento.

Putnam defende um ramo do funcionalismo hoje chamado de *funcionalismo de máquina*. Ele enfatiza **autômatos probabilísticos**, os quais são similares às máquinas de Turing, exceto que as transições entre estados computacionais são estocásticas. Ele propõe que a atividade mental implementa um autômato probabilístico e que estados mentais particulares são estados de máquina do processador central desse autômato. A tabela de máquina especifica a organização funcional apropriada e também especifica o papel que os estados mentais individuais desempenham dentro dessa organização funcional. Desse modo, Putnam combina o funcionalismo com a TCCM.

O funcionalismo de máquina enfrenta muitos problemas. Um problema, enfatizado por Ned Block e Jerry Fodor (1972), diz respeito à **produtividade do pensamento**. Um humano normal pode entreter uma potencial infinidade de proposições. O funcionalismo de máquina identifica estados mentais com estados de máquina de um autômato probabilístico. Já que há apenas uma quantidade finita de estados de máquina, não há estados de máquina suficientes para parear um a um com os possíveis estados mentais de um ser humano normal. Obviamente, um humano real sempre entreterá apenas uma quantidade finita de proposições. No entanto, Block e Fodor afirmam que essa limitação reflete limites de memória e tempo de vida, ao invés de (digamos) alguma lei psicológica que restringe a classe das proposições possíveis de serem entretidas por humanos. Um autômato probabilístico é dotado de tempo e capacidade de memória infinitos, e ainda assim tem apenas uma quantidade finita de estados de máquina. Aparentemente, então, o funcionalismo de máquina atribui mal os limites finitários à cognição humana.

Outro problema para o funcionalismo de máquina, também enfatizado por Block e Fodor (1972), diz respeito à **sistematicidade do pensamento**. Uma habilidade de entreter uma proposição se correlaciona com uma habilidade de pensar outras proposições. Por exemplo, alguém que pode entreter o pensamento de que **João ama Maria** também pode entreter o pensamento de que **Maria ama**

**João.** Portanto, parece haver relações sistemáticas entre estados mentais. Uma boa teoria deveria refletir essas relações sistemáticas. No entanto, o funcionalismo de máquina identifica estados mentais com estados de máquina não estruturados, os quais carecem das requisitadas relações sistemáticas com os demais. Por essa razão, o funcionalismo de máquina não explica a sistematicidade. Em resposta a essa objeção, funcionalistas de máquina podem negar que eles sejam obrigados a explicar a sistematicidade. Ainda assim, a objeção sugere que o funcionalismo de máquina negligencia características essenciais da mentalidade humana. Uma teoria melhor explicaria essas características a partir de princípios.

Embora as objeções da produtividade e da sistematicidade ao funcionalismo de máquina possam não ser decisivas, elas fornecem um forte ímpeto para buscar uma versão melhorada da TCCM. Veja Block (1978) para problemas adicionais enfrentados pelo funcionalismo de máquina e pelo funcionalismo em geral.

### 3.2 A teoria representacional da mente

Fodor (1975, 1981, 1987, 1990, 1994, 2008) defende uma versão da TCCM que acomoda a sistematicidade e a produtividade de forma muito mais satisfatória. Ele desloca a atenção para os **símbolos** manipulados durante a computação ao estilo Turing.

Uma visão antiga, que remonta até, pelo menos, a *Summa Logicae*<sup>7</sup> de Guilherme de Ockham, sustenta que o pensamento ocorre em uma *linguagem do pensamento* (algumas vezes chamada de *mentalês* [*Mentalese*]). Fodor revive essa visão. Ele postula um sistema de representações mentais, incluindo tanto representações primitivas quanto representações complexas formadas de representações primitivas. Por exemplo, as palavras primitivas em mentalês **João**, **Maria** e **ama** podem ser combinadas para formar a frase complexa em mentalês **João ama Maria**. O mentalês é *composicional*: o significado de uma expressão complexa em mentalês é uma função dos significados das suas partes e do modo como essas partes são combinadas. Atitudes proposicionais são relações com símbolos em mentalês. Fodor chama essa visão de *a teoria representacional da mente* (TRM). Combinando a TRM com a TCCM, ele argumenta que a atividade mental envolve computação

---

<sup>7</sup> N.T.: *Suma de Lógica*.

ao estilo Turing sobre a linguagem do pensamento. A computação mental armazena símbolos em mentalês em locais de memória, manipulando esses símbolos de acordo com regras mecânicas.

Uma das principais virtudes da TRM é a rapidez com que acomoda a produtividade e a sistematicidade:

*Produtividade:* a TRM postula um conjunto finito de expressões em mentalês, combináveis em potencialmente infinitas expressões complexas em mentalês. Uma pensante com acesso ao vocabulário mentalês primitivo e aos dispositivos de composição do mentalês tem o potencial para entreter uma quantidade infinita de expressões em mentalês. Ela, então, tem o potencial para instanciar uma quantidade infinita de atitudes proposicionais (desconsiderando os limites de tempo e memória).

*Sistematicidade:* de acordo com a TRM, há relações sistemáticas entre as atitudes proposicionais que uma pensante pode entreter. Por exemplo, suponha que eu posso pensar que João ama Maria. De acordo com a TRM, esse meu pensar envolve que eu esteja em uma relação  $R$  com uma frase em mentalês **João ama Maria**, composta das palavras em mentalês **João**, **ama** e **Maria** combinadas do modo certo. Se em tenho essa capacidade, então eu também tenho a capacidade de estar em uma relação  $R$  com a frase em mentalês distinta **Maria ama João**, e, assim, estar pensando que Maria ama João. Portanto, a capacidade de pensar que João ama Maria é sistematicamente relacionada com a capacidade de pensar que Maria ama João.

Ao tratar as atitudes proposicionais como relações com símbolos mentais complexos, a TRM explica tanto a produtividade quanto a sistematicidade.

A TCCM+TRM difere do funcionalismo de máquina em vários outros pontos. Primeiramente, o funcionalismo de máquina é uma teoria sobre estados mentais *em geral*, enquanto a TRM é uma teoria apenas acerca de atitudes proposicionais. Em segundo lugar, proponentes da TCCM+TRM não precisam dizer que atitudes proposicionais são funcionalmente individuadas. Como aponta Fodor (2000, p. 105), precisamos distinguir o *computacionalismo* (processos mentais são computacionais) do *funcionalismo* (estados mentais são estados funcionais). O funcionalismo de máquina endossa ambas as teses. A TCCM+TRM endossa apenas a primeira. Infelizmente, muitos filósofos ainda presumem erroneamente que o computacionalismo implica uma abordagem funcionalista sobre as atitudes proposicionais (*vide* PICCININI, 2004).

Discussões filosóficas sobre a TRM tendem a focar, majoritariamente, em **pensamento humano de alto nível**, especialmente crenças e desejos. No entanto, a TCCM+TRM é aplicável a um conjunto mais amplo de estados e processos

mentais. Muitos cientistas cognitivos a aplicam a animais não humanos. Por exemplo, Gallistel e King (2009) a aplicam a certos fenômenos de invertebrados (por exemplo, navegação de abelhas). Mesmo quando se restringe a análise a humanos, pode-se aplicar a TCCM+TRM a *processamentos subpessoais*. Fodor (1983) argumenta que a percepção envolve um “módulo” subpessoal que converte entradas retiniais em símbolos em mentalês e então executa computação sobre esses símbolos. Portanto, falar sobre uma linguagem do *pensamento* é potencialmente enganoso, já que isso sugere uma inexistente restrição a atividades mentais de alto nível.

Também potencialmente enganosa é a descrição do mentalês como uma **linguagem**, o que sugere que todos os símbolos mentalês se assemelham a expressões em linguagem natural. Muitos filósofos, incluído Fodor, aparentam endossar essa posição às vezes. No entanto, há formatos não proposicionais possíveis para os símbolos do mentalês. Proponentes da TCCM+TRM podem adotar uma linha pluralista, permitindo que a computação mental opere sobre itens semelhantes a imagens, mapas, diagramas, ou outras representações não proposicionais (JOHNSON-LAIRD, 2004, p. 187; MCDERMOTT, 2001, p. 69; PINKER, 2005, p. 7; SLOMAN, 1978, p. 144-176). A linha pluralista parece especialmente plausível quando aplicada a processos subpessoais (tais como a percepção) e a animais não humanos. Michael Rescorla (2009a; 2009b) analisa pesquisas sobre *mapas cognitivos* (TOLMAN, 1948; O’KEEFE; NADEL, 1978; GALLISTEL, 1990), sugerindo que alguns animais podem navegar pelo ambiente ao executar computações sobre representações mentais mais semelhantes a mapas do que a frases. Elisabeth Camp (2009), citando pesquisas sobre interações sociais de babuínos (CHENEY; SEYFARTH, 2007), argumenta que babuínos podem codificar relações de dominância social por meio de representações que não são frasais nem estruturadas em formato de árvore.

A TCCM+TRM é esquemática. Para preencher o esquema, deve-se prover modelos computacionais detalhados dos processos mentais específicos. Um modelo completo irá:

- descrever os símbolos em mentalês manipulados pelo processo;
- isolar operações elementares que manipulam os símbolos (por exemplo, **inscrever um símbolo em um local de memória**); e
- delinear regras mecânicas que governam a aplicação das operações elementares.

Ao prover um modelo computacional detalhado, decomparamos um processo mental complexo em uma série de operações elementares governadas por instruções de rotina precisas.

A TCCM+TRM permanece neutra com respeito ao debate tradicional entre o fisicismo e o dualismo de substância. Um modelo ao estilo Turing opera em um nível muito abstrato, sem dizer se as computações mentais são implementadas em coisas físicas ou em coisas como uma alma nos moldes cartesianos (BLOCK, 1983, p. 522). Na prática, todos os proponentes da TCCM+TRM abraçam um modelo fisicista amplo. Eles defendem que computações mentais são implementadas não por coisas como almas, mas, sim, pelo cérebro. Nessa visão, símbolos em mentalês são realizados por estados neurais e operações sobre símbolos em mentalês são realizadas por processos neurais. Em última análise, proponentes fisicistas da TCCM+TRM precisam produzir teorias empíricas amplamente confirmadas que expliquem como exatamente a atividade neural implementa uma computação ao estilo Turing. Como Gallistel e King (2009) enfatizam, atualmente não temos tais teorias, confira, no entanto, Zylberberg, Dehaene, Roelfsema e Sigman (2011) para algumas especulações.

Fodor (1975) desenvolve a TCCM+TRM como uma base para as ciências cognitivas. Ele discute fenômenos mentais tais como tomada de decisão, percepção e processamento de linguagem. Em cada caso, ele sustenta, as nossas melhores teorias científicas postulam computação ao estilo Turing sobre representações mentais. De fato, ele argumenta que nossas **únicas** teorias viáveis têm essa forma. Ele conclui que a TCCM+TRM é a única opção disponível. Muitos cientistas cognitivos argumentam nessa mesma linha. C. R. Gallistel e Adam King (2009), Philip Johnson-Laird (1988), Allen Newell e Herbert Simon (1976) e Zenon Pylyshyn (1984), todos tomam a computação ao estilo Turing sobre símbolos mentais como a melhor base para uma teorização científica a respeito da mente.

#### 4. Redes neurais

Na década de 1980, o connexionismo emergiu como um rival proeminente ao computacionalismo clássico. Conexionistas inspiram-se na neurofisiologia ao invés de na lógica e na ciência da computação. Eles empregam modelos computacionais,



**redes neurais**, que diferem significativamente dos modelos ao estilo Turing. Uma **rede neural** é uma coleção de nós interconectados. Nós caem sob três categorias: nós **de entrada**, nós **de saída** e nós **ocultos** (os quais mediam os nós de entrada e de saída). Os nós têm valores de ativação (*activation values*), dados por números reais. Um nó pode manter uma *conexão ponderada* (*weighted connection*) com outro nó, a qual também é dada por um número real. As ativações dos nós de entrada são determinadas exogenamente: essas são as entradas da computação. A **ativação total de entrada** de um nó oculto ou de saída é uma soma ponderada da ativação dos nós que a alimentam. A ativação de um nó oculto ou de saída é uma função de sua ativação total de entrada; a função particular varia de acordo com a rede. Durante a computação numa rede neural, ondas de ativação se propagam dos nós de entrada até os nós de saída, determinadas pelas conexões ponderadas entre os nós.

Em uma **rede não circular** (*feedforward network*), as conexões ponderadas fluem em apenas uma direção. *Redes recorrentes* (*recurrent networks*) têm laços de retorno (*feedback loops*), nos quais conexões emanadas de unidades ocultas retornam para unidades ocultas. Redes recorrentes são menos tratáveis matematicamente do que as redes não circulares. No entanto, elas são cruciais na modelagem psicológica de vários fenômenos, tais como os fenômenos que envolvem algum tipo de memória (ELMAN, 1990).

Os pesos (*weights*) em uma rede neural são tipicamente mutáveis, evoluindo de acordo com um **algoritmo de aprendizagem**. A literatura oferece vários algoritmos de aprendizagem, mas a ideia básica normalmente é a de ajustar os pesos para que as **saídas reais** gradualmente se aproximem das *saídas desejadas* as quais são esperadas a partir das entradas relevantes. O **algoritmo de retropropagação** (*backpropagation algorithm*) é um algoritmo desse tipo amplamente utilizado (RUMELHART; HINTON; WILLIAMS, 1986).

O conexionismo remonta a McCulloch e Pitts (1943), que estudaram redes de interconexão de *portas lógicas* (por exemplo, porta-E e porta-OU). Pode-se ver uma rede de portas lógicas como uma rede neural, com ativações restritas a dois valores (0 e 1) e funções de ativação dadas pelas funções de verdade habituais. McCulloch e Pitts desenvolveram as portas lógicas como modelos idealizados de neurônios individuais. A discussão que propuseram exerceu uma profunda influência da ciência da computação (VON NEUMANN, 1945). Computadores digitais modernos são simplesmente redes de portas lógicas. Dentro das ciências cognitivas, no entanto, pesquisadores geralmente focam em redes cujos elementos sejam mais “parecidos com neurônios” do que as portas lógicas. Mais especificamente,

conexionistas nos dias de hoje geralmente enfatizam as redes neurais analógicas, cujos nós têm valores contínuos ao invés de discretos. Alguns autores chegam a utilizar a expressão “redes neurais” como denotando exclusivamente redes desse último tipo.

As redes neurais receberam relativamente pouca atenção dos cientistas cognitivos durante as décadas de 1960 e 1970, quando os modelos ao estilo Turing dominavam. Mas a década de 1980 testemunhou um grande ressurgimento do interesse em redes neurais, especialmente redes neurais analógicas, com os dois volumes de *Parallel Distributed Processing*<sup>8</sup> (RUMELHART; MCCLELLAND; THE PDP RESEARCH GROUP, 1986; MCCLELLAND, RUMELHART; THE PDP RESEARCH GROUP, 1987) servindo como um manifesto. Pesquisadores construíram modelos connexionistas de diversos fenômenos: reconhecimento de objetos, percepção de fala, compreensão de frases, desenvolvimento cognitivo e assim por diante. Impressionados com o connexionismo, muitos pesquisadores concluíram que a TCCM+TRM não era mais a única opção disponível.

Na década de 2010, uma classe de modelos computacionais conhecidos como *redes neurais profundas* (*deep neural networks*) tornou-se bastante popular (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; LECUN; BENGIO; HINTON, 2015). Esses modelos são redes neurais com múltiplas camadas de nós ocultos (às vezes centenas de tais camadas). Redes neurais profundas - treinadas em conjuntos imensos de dados a partir de um ou outro algoritmo de aprendizagem (geralmente de retropropagação) - atingiram um grande sucesso em muitas áreas da IA, incluindo reconhecimento de objetos e jogos de estratégia. Redes neurais profundas são amplamente empregadas em aplicações comerciais e são foco de extensas investigações em andamento tanto na academia quanto na indústria. Pesquisadores começaram a utilizá-las também para modelar a mente (por exemplo, MARBLESTONE, WAYNE; KORDING, 2016; KRIEGESKORTE, 2015).

Para uma visão detalhada sobre redes neurais, veja Haykin (2008). Para uma introdução amigável, com ênfase nas aplicações psicológicas, veja Marcus (2001). Para uma introdução filosoficamente orientada de redes neurais profundas, veja Buckner (2019).

---

<sup>8</sup> N.T.: *Processamento Distribuído Paralelo*.

## 4.1 Relações entre a computação das redes neurais e a computação clássica

As redes neurais têm uma “pegada” muito diferente dos modelos clássicos (ou seja, ao estilo Turing). Ainda assim, a computação clássica e a computação das redes neurais não são mutualmente excludentes:

- *Pode-se implementar uma rede neural em um modelo clássico.* De fato, toda rede neural já construída fisicamente foi implementada em um computador digital.
- *Pode-se implementar um modelo clássico em uma rede neural.* Os computadores digitais modernos implementam computação ao estilo Turing em redes de portas lógicas. Alternativamente, pode-se implementar computação ao estilo Turing utilizando uma rede neural recorrente analógica, cujos nós têm valores contínuos de ativação (SIEGELMANN; SONTAG, 1991; SIEGELMANN; SONTAG, 1995)<sup>9</sup>.

Embora alguns pesquisadores sugiram uma oposição fundamental entre a computação clássica e a das redes neurais, parece mais preciso identificar duas tradições de modelagem que em alguns casos se sobrepõem e em outros não (*vide* BODEN, 1991; PICCININI, 2008b). Nessa linha, é também digno de nota que tanto o computacionalismo clássico quanto o computacionalismo conexionista têm sua origem comum no trabalho de McCulloch e Pitts.

Filósofos frequentemente dizem que a computação clássica envolve “manipulação de símbolos governada por regras” enquanto a computação realizada pelas redes neurais é não simbólica. O quadro intuitivo é o de que a “informação” nas redes neurais é distribuída globalmente ao longo dos pesos e das ativações, ao invés de concentrada em símbolos localizados. No entanto, a noção de “símbolo” ela mesma requer explicações, já que geralmente não é claro o que os teóricos querem dizer ao descrever a computação como simbólica *versus* não simbólica. Como mencionado na seção 1, o formalismo de Turing impõe muito poucas condições

---

<sup>9</sup>N.T.: O autor também inclui como fonte o texto *Neural Turing Machines*, manuscrito de Graves, Wayne e Danihelka (2014). Disponível em: <https://arxiv.org/abs/1410.5401>. Acesso em: 20 jan. 2022.

aos “símbolos”. Com respeito a símbolos primitivos, Turing assume apenas que há uma quantidade finita deles e que podem ser inscritos em locais de memória de leitura/gravação. Redes neurais também podem manipular símbolos que satisfazem essas duas condições: como já dito, pode-se implementar um modelo ao estilo Turing em uma rede neural.

Muitas discussões sobre a dicotomia simbólica/não-simbólica empregam uma noção mais robusta de “símbolo”. Na abordagem mais robusta, um símbolo é o tipo de coisa que representa algo. Assim, uma coisa é um símbolo apenas se tem propriedades semânticas ou representacionais. Se empregarmos essa noção mais robusta de símbolo, então a distinção simbólica/não-simbólica corta verticalmente a distinção entre computação ao estilo Turing e a computação das redes neurais. Uma máquina de Turing não precisa empregar símbolos no sentido mais robusto. Até onde vai o formalismo de Turing, símbolos manipulados durante a computação de Turing não precisam ter propriedades representacionais (CHALMERS, 2011). Por outro lado, uma rede neural pode manipular símbolos com propriedades representacionais. Na verdade, uma rede neural analógica pode manipular símbolos que tenham uma sintaxe combinatória e uma semântica (HORGAN; TIENSON, 1996; MARCUS, 2001).

Seguindo Steven Pinker e Alan Prince (1988), podemos distinguir entre o *conexionismo eliminativo* (*eliminative connectionism*) e o *conexionismo implementacionista* (*implementationist connectionism*).

Conexionistas eliminativos desenvolvem o conexionismo como um rival do computacionalismo clássico. Eles argumentam que o formalismo de Turing é irrelevante para a explicação psicológica. Geralmente, embora nem sempre, eles tentam reviver a tradição *associacionista* (*associationist*) na psicologia, uma tradição que a TCCM fortemente desafiou. Geralmente, embora nem sempre, eles atacam a linguística mentalista e inatista lançada por Noam Chomsky (1965). Geralmente, embora nem sempre, eles manifestam notória hostilidade à própria noção de representação mental. Mas a característica definidora do conexionismo eliminativo é que ele utiliza as redes neurais como **substitutos** aos modelos ao estilo Turing. Conexionistas eliminativos veem a mente como um sistema computacional de um tipo radicalmente diferente do da máquina de Turing. Alguns autores defendem o conexionismo eliminativo explicitamente (CHURCHLAND, 1989; RUMELHART; MCCLELLAND, 1986; HORGAN; TIENSON, 1996), e muitos outros se inclinam para ele.

O conexionismo implementacionista é uma posição mais ecumênica. Ela permite um papel potencialmente valioso **tanto** para os modelos ao estilo Turing

*quanto* para as redes neurais, operando harmoniosamente em níveis diferentes de descrição (MARCUS, 2001; SMOLENSKY, 1988). Um modelo ao estilo Turing é um modelo de nível mais alto, enquanto um modelo de rede neural é um modelo de nível mais baixo. A rede neural lança luz sobre como o cérebro implementa o modelo ao estilo Turing, exatamente como uma descrição em termos de portas lógicas lança luz acerca de como um computador pessoal executa um programa escrito em uma linguagem de programação de alto nível.

## 4.2 Argumentos a favor do conexionismo

O conexionismo entusiasma muitos pesquisadores por causa da analogia entre redes neurais e o cérebro. Nós se parecem com neurônios, enquanto as conexões entre os nós se parecem com sinapses. A modelagem conexionista parece, portanto, mais “biologicamente plausível” que a modelagem clássica. Um modelo conexionista de um fenômeno psicológico aparentemente captura (de um modo ideal) como neurônios interconectados podem gerar o fenômeno.

Ao avaliar o argumento da plausibilidade biológica, deve-se reconhecer que redes neurais variam amplamente a respeito do quão bem elas se parecem com a atividade real do cérebro. Muitas redes que se destacam nos textos conexionistas não são tão biologicamente plausíveis (BECHTEL; ABRAHAMSEN, 2002, p. 341-343; BERMÚDEZ, 2010, p. 237-239; CLARK, 2014, p. 87-89; HARNISH, 2002, p. 359-362). Alguns exemplos:

- Neurônios reais são muito mais heterogêneos do que os nós intersubstituíveis que figuram em redes conexionistas típicas.
- Neurônios reais emitem disparos (*spikes*) discretos (potenciais de ação [*action potentials*]) como saídas. Mas os nós que figuram em muitas redes neurais proeminentes, incluindo as redes neurais profundas mais conhecidas, ao contrário, têm saídas contínuas.
- O algoritmo de retropropagação requer que os pesos (*weights*) entre os nós possam variar entre excitador (*excitatory*) e inibidor (*inhibitory*), no entanto, as sinapses reais não podem variar assim (CRICK; ASANUMA, 1986). Além disso, o algoritmo assume

saídas desejadas suplementadas exogenamente por modeladores **que sabem a resposta desejada**. Nesse sentido, a aprendizagem é **supervisionada**. Muito pouca aprendizagem nos sistemas biológicos reais envolve qualquer coisa parecida com treinamento supervisionado.

Por outro lado, algumas redes neurais são mais biologicamente realistas (BUCKNER; GARSON, 2019; ILLING; GERSTNER; BREA, 2019). Por exemplo, há redes neurais que substituem a retropropagação por algoritmos de aprendizagem mais realistas, como um algoritmo de aprendizagem por reforço (*reinforcement learning algorithm*)<sup>10</sup> ou um algoritmo de aprendizagem não supervisionada (KROTOV; HOPFIELD, 2019). Há também redes neurais cujos nós emitem disparos discretos grosso modo parecidos com os emitidos por neurônios reais no cérebro (MAASS, 1996; BUESING; BILL; NESSLER; MAASS, 2011).

Mesmo quando uma rede neural não é biologicamente plausível, ela ainda é *mais* biologicamente plausível que modelos clássicos. Redes neurais certamente parecem mais próximas, tanto em detalhes quanto em “espírito”, de descrições neurofisiológicas do que modelos ao estilo Turing. Muitos cientistas cognitivos temem que a TCCM reflita uma tentativa errônea de impor a arquitetura de computadores digitais ao cérebro. Alguns duvidam que o cérebro implemente qualquer coisa parecida com a computação digital, ou seja, computação sobre configurações discretas de dígitos (PICCININI; BAHAR, 2013). Outros duvidam que o cérebro exiba uma separação clara ao estilo Turing entre o processador central e a memória de leitura/gravação (DAYAN, 2009). Redes neurais se saem melhor em ambas as alas: não requerem computação sobre configurações discretas de dígitos e não postulam uma separação clara entre processador central e memória de leitura/gravação.

Computacionalistas clássicos tipicamente respondem que é prematuro tirar conclusões baseadas em plausibilidade biológica, dado o quão pouco entendemos sobre a relação entre descrições ao nível neural, computacional e cognitivo (GALLISTEL; KING, 2009; MARCUS, 2001). Utilizando técnicas de mensuração

---

<sup>10</sup> N. T.: O autor recomenda como fonte o manuscrito *A Biologically Plausible Learning Rule for Deep Learning in the Brain*, de Isabella Pozzi, Sander M. Bohtë e Pieter R. Roelfsema (2019). Disponível em: <https://arxiv.org/pdf/1811.01768.pdf>. Acesso em: 22 jan. 2022.

tais como a gravação celular (*cell recordings*) e a imagem por ressonância magnética funcional (fMRI - *functional magnetic resonance imaging*, em inglês) e se apoiando em disciplinas tão diversas quanto a física, a biologia, a IA, a teoria da informação (*information theory*), a estatística, a teoria dos grafos (*graph theory*) e a teoria dos sistemas dinâmicos, neurocientistas têm acumulado um conhecimento substancial sobre o cérebro em vários níveis de refinamento (ZEDNIK, 2019). Hoje sabemos muito sobre neurônios individuais, sobre como neurônios interagem dentro de uma população neural, sobre a localização de atividade mental em regiões corticais (por exemplo, o córtex visual) e sobre interações entre regiões corticais. Ainda assim, temos imensamente muito a aprender sobre como o tecido neural realiza as tarefas que certamente realiza: percepção, raciocínio, tomada de decisão, aquisição da fala e tantas outras. Dado o nosso atual estado de relativa ignorância, seria precipitado insistir que o cérebro não implementa qualquer coisa parecida com computação ao estilo Turing.

Conexionistas oferecem muitos outros argumentos para sustentar que deveríamos empregar modelos conexionistas ao invés de, ou paralelamente a, modelos clássicos. Veja o verbete da SEP, sobre **connectionism**, para uma visão geral. Para os propósitos deste verbete, mencionaremos mais dois argumentos.

O primeiro argumento enfatiza a **aprendizagem** (BECHTEL; ABRAHAMSEN, 2002, p. 51). Um grande número de fenômenos cognitivos envolve aprendizagem a partir da experiência. Muitos modelos conexionistas são explicitamente desenhados para modelar a aprendizagem por meio de retropropagação ou algum outro algoritmo que modifique o peso entre os nós. Por outro lado, conexionistas frequentemente reclamam que não há bons modelos clássicos para a aprendizagem. Computacionalistas clássicos podem responder citando defeitos constatáveis nos algoritmos de aprendizagem conexionistas (por exemplo, a forte dependência de retropropagação em treinamento supervisionado). Computacionalistas clássicos também podem mencionar a teoria da decisão bayesiana, que modela a aprendizagem como atualização probabilística. Mais especificamente, computacionalistas clássicos podem mencionar as façanhas das *ciências cognitivas bayesianas*, que utilizam a teoria da decisão bayesiana para construir modelos matemáticos da atividade mental (MA, 2019). No curso das últimas décadas, as ciências cognitivas bayesianas acumularam muitos sucessos explicativos. Esse histórico impressionante sugere que alguns processos mentais são bayesianos ou aproximadamente bayesianos (RESCORLA, 2020). Além disso, os avanços mencionados na seção 2 mostram como sistemas computacionais clássicos podem executar, ou pelo menos executar de forma aproximada, atualizações bayesianas em vários cenários realistas. Esses

desenvolvimentos fornecem esperanças de que a computação clássica possa modelar diversos casos importantes de aprendizagem.

O segundo argumento enfatiza a *velocidade da computação*. Neurônios são muito mais lentos que componentes à base de silício dos computadores digitais. Por essa razão, neurônios não poderiam executar uma computação seriada com rapidez suficiente para corresponder à veloz performance humana na percepção, compreensão de fala, tomada de decisão, etc. Conexionistas sustentam que a única solução viável é substituir a computação seriada por uma arquitetura computacional “massivamente paralela”, exatamente o que as redes neurais fazem (FELDMAN; BALLARD, 1982; RUMELHART, 1989). No entanto, esse argumento é efetivo apenas contra computacionalistas clássicos que insistem em processamento seriado. Como dito na seção 3, alguns modelos ao estilo Turing envolvem processamento paralelo. Muitos computacionalistas clássicos aceitam felizes uma computação mental “massivamente paralela” e o argumento não tem efeito contra esses pesquisadores. Dito isso, o argumento explicita uma questão importante que qualquer computacionalista, seja clássico, conexionista ou qualquer outro, precisa responder: Como um cérebro construído de neurônios relativamente lentos executa computações sofisticadas com tanta velocidade? Nem computacionalistas clássicos nem conexionistas responderam satisfatoriamente a essa questão (GALLISTEL; KING, 2009, p. 174; p. 265).

### 4.3 Sistemática e produtividade

Fodor e Pylyshyn (1988) oferecem uma crítica ao conexionismo eliminativista amplamente discutida. Eles argumentam que a sistematicidade e a produtividade não ocorrem nos modelos conexionistas, exceto quando o modelo conexionista implementa um modelo clássico. Portanto, o conexionismo não constitui uma alternativa viável à TCCM. Na melhor das hipóteses, fornece uma descrição de baixo-nível que ajuda a preencher a lacuna entre a computação ao estilo Turing e a descrição neurocientífica.

Essa argumentação fomentou muitas respostas e contrarrespostas. Alguns argumentam que as redes neurais podem exibir sistematicidade sem implementar qualquer coisa parecida com a arquitetura computacional clássica (HORGAN; TIENSON, 1996; CHALMERS, 1990; SMOLENSKY, 1991; VAN GELDER, 1990). Alguns argumentam que Fodor e Pylyshyn exageraram demasiadamente a sistematicidade (JOHNSON, 2004) ou a produtividade (RUMELHART; MCCLELLAND



1986), especialmente para animais não humanos (DENNETT, 1991). Tais problemas, assim como muitos outros levantados pela argumentação de Fodor e Pylyshyn, têm sido amplamente investigados na literatura. Para uma maior discussão, veja Bechtel e Abrahamsen (2002, p. 156-199), Bermúdez (2005, p. 244-278), Chalmers (1993), Clark (2014, p. 84-86) e os verbetes da SEP **the language of thought hypothesis** e **connectionism**.

Gallistel e King (2009) desenvolveram um argumento da produtividade parecido, mas distinto. Eles enfatizaram a *produtividade da computação mental*, em oposição à *produtividade dos estados mentais*. A partir de detalhados estudos de casos empíricos, argumentaram que muitos animais não humanos podem extrair, armazenar e recuperar registros detalhados do ambiente que os circula. Por exemplo, o pássaro *Aphelocoma californica*<sup>11</sup> registra onde estocou alimento, qual tipo de alimento estocou em cada localização, quando estocou o alimento e se esgotou os recursos de um determinado estoque (CLAYTON; EMERY; DICKINSON, 2006). Esse pássaro pode acessar esses registros e utilizá-los em diversas computações: computar se um alimento armazenado em um certo estoque pode ter se deteriorado; computar uma rota de uma localização para outra e assim por diante. O número de computações que esse pássaro pode executar é, para todos os efeitos práticos, infinito.

A TCCM explica a produtividade da computação mental postulando um processador central que armazena e recupera símbolos em memória endereçável de gravação/leitura. Quando precisa, o processador central pode recuperar combinações arbitrárias e imprevistas de símbolos da memória. Gallistel e King argumentam que o conexionismo, ao contrário, tem dificuldade em acomodar a produtividade da computação mental. Embora Gallistel e King não façam uma distinção cuidadosa entre o conexionismo eliminativista e o implementacionista, podemos resumir o argumento deles da seguinte forma:

- O conexionismo eliminativista não consegue explicar como os organismos combinam memórias armazenadas (por exemplo, em locais de estocagem) para fins computacionais (por exemplo, computar uma rota de um estoque para outro). Há virtualmente uma infinidade de possibilidades de combinação que podem ser

---

<sup>11</sup> N.T.: *Western Scrub Jay*, no original.

úteis, sem previsão antecipada sobre quais peças de informação devem ser combinadas por computações futuras. A única solução computacionalmente tratável é o armazenamento de símbolos em locais de memória de gravação/leitura prontamente acessíveis - uma solução que os conexionistas eliminativistas rejeitam.

- Conexionistas implementacionistas podem postular o armazenamento de símbolos numa memória de leitura/gravação, *como implementada por uma rede neural*. No entanto, os mecanismos que os conexionistas geralmente propõem para a implementação de memória não são plausíveis. As propostas existentes são majoritariamente variações da mesma ideia: uma rede neural recorrente que permite que a atividade de reverberação viaje ao longo de um laço (ELMAN, 1990). Há muitas razões para sustentar que o modelo de laço reverberatório é desanimador enquanto uma teoria acerca da memória de longo prazo (*long-term memory*). Por exemplo, ruído no sistema nervoso assegura que os sinais iriam rapidamente se degradar em poucos minutos. Conexionistas implementacionistas até agora não ofereceram um modelo plausível para a memória de gravação/leitura<sup>12</sup>.

Gallistel e King concluem que a TCCM é muito mais adequada para explicar um vasto conjunto de fenômenos cognitivos do que o conexionismo eliminativista e o implementacionista.

---

<sup>12</sup> A ciência da computação oferece muitas técnicas para a implementação da memória de leitura/gravação nas redes neurais. Por exemplo, se usarmos uma rede neural recorrente analógica adequada, então poderemos codificar os conteúdos da fita de memória nos valores de ativação dos nós (SIEGELMANN; SONTAG, 1995). No entanto, conexionistas implementacionistas não propõem que a memória nos sistemas biológicos trabalhe, na realidade, desse modo, talvez porque eles considerem a implementação biologicamente implausível (HADLEY, 2000).

Os críticos atacam esse novo argumento da produtividade por vários ângulos, focando majoritariamente nos estudos de casos empíricos apresentados por Gallistel e King. Peter Dayan (2009), John Donahoe (2010) e Christopher Mole (2014) argumentam que modelos de redes neurais biologicamente plausíveis podem acomodar pelo menos alguns dos casos estudados. Dayan e Donahoe argumentam que modelos de redes neurais empiricamente adequados podem prescindir de qualquer coisa que pareça uma memória de leitura/gravação. Mole argumenta que, em certos casos, modelos de redes neurais empiricamente adequados podem **implementar** os mecanismos de memória de leitura/gravação postos por Gallistel e King. O debate desses problemas fundamentais parece destinado a continuar no futuro.

#### 4.4 Neurociência computacional

A *neurociência computacional* descreve o sistema nervoso por meio de modelos computacionais. Embora esse programa de pesquisa seja ancorado em modelagem matemática de neurônios individuais, o foco distintivo da neurociência computacional é **sistemas** de neurônios interconectados. A neurociência computacional geralmente modela esses sistemas como redes neurais. Nesse sentido, ela é uma variante, uma ramificação ou uma descendente do conexionismo. No entanto, a maioria dos neurocientistas computacionais não se autointitulam conexionistas. Há várias diferenças entre o conexionismo e a neurociência computacional:

- As redes neurais empregadas pelos neurocientistas computacionais são biologicamente muito mais realistas que as empregadas pelos conexionistas. A literatura neurocientífica computacional é repleta de falas sobre taxas de disparo (*firing rates*), potenciais de ação (*action potentials*), curvas de sincronização (*tuning curves*), etc. Essas noções desempenham um papel, na melhor das hipóteses, limitado nas pesquisas conexionistas, tal qual a maioria das pesquisas levantadas em Rogers e McClelland (2014).
- A neurociência computacional é impulsionada em grande medida por conhecimento sobre o cérebro e atribui enorme importância aos dados neurofisiológicos (por exemplo, gravação celular). Conexionistas

atribuem muito menos ênfase a esses dados. Sua pesquisa é impulsionada primariamente por dados comportamentais (embora textos conexionistas mais recentes cite dados neurofisiológicos com uma frequência um pouco maior).

- Neurocientistas computacionais geralmente tomam os nós individuais nas redes neurais como descrições ideais de neurônios reais. Conexionistas, ao contrário, geralmente tomam os nós como *unidades de processamento parecidas com neurônios* (ROGERS; MCCLELLAND, 2014), permanecendo neutros sobre como exatamente essas unidades mapeiam entidades neurofisiológicas reais.

Pode-se dizer que a neurociência computacional está preocupada majoritariamente com a *computação neural* (computação realizada por um sistema de neurônios), enquanto o conexionismo está preocupado majoritariamente com modelos computacionais abstratos **inspirados** por computação neural. Mas as fronteiras entre o conexionismo e a neurociência computacional são admitidamente um tanto porosas. Para uma visão geral da neurociência computacional, consulte Trappenberg (2010) ou Miller (2018).

O engajamento sério da filosofia com as neurociências remonta pelo menos ao *Neurophilosophy* (1986)<sup>13</sup> de Patricia Churchland. À medida que as neurociências amadureceram, Churchland se tornou um de seus defensores filosóficos mais importantes (CHURCHLAND; KOCH; SEJNOWSKI, 1990; CHURCHLAND; SEJNOWSKI, 1992). Se juntaram a ela Paul Churchland (1995; 2007) e outros (ELIASMITH, 2013; ELIASMITH; ANDERSON, 2003; PICCININI; BAHAR, 2013; PICCININI; SHAGRIR, 2014). Todos esses autores sustentam que a teorização sobre a computação mental deveria começar com o cérebro, não com máquinas de Turing ou outras ferramentas inapropriadas extraídas da lógica e da ciência da computação. Eles também defendem que a modelagem de redes neurais deveria se esforçar para obter um realismo biológico maior do que aquele tipicamente obtido pelos modelos conexionistas. Chris Eliasmith (2013) desenvolve esse ponto de vista neurocomputacional por meio do *Enquadramento da Engenharia Neural* [*Neural*

---

<sup>13</sup> N.T.: *Neurofilosofia*.

*Engineering Framework*], o qual suplementa a neurociência computacional com ferramentas oriundas da teoria do controle (BROGAN, 1990). Ele pretende produzir uma “engenharia reversa” do cérebro, construindo modelos de redes neurais de larga escala e biologicamente plausíveis para fenômenos cognitivos.

A neurociência computacional difere da TCCM e do conexionismo em um ponto crucial: ela abandona a múltipla realizabilidade. Neurocientistas computacionais citam propriedades e processos neurofisiológicos específicos, assim, seus modelos não se aplicam igualmente bem a (digamos) uma criatura suficientemente diferente à base de silício. Desse modo, a neurociência computacional sacrifica uma característica central que originalmente fez com que os filósofos se atraíssem pela TCM. Os neurocientistas computacionais responderão que as descobertas resultantes sobre os fundamentos neurofisiológicos valem esse sacrifício. Mas muitos computacionalistas temem que, ao focar de mais em fundamentos neurais, corremos o risco de perder de vista a floresta cognitiva para as árvores neurais. Detalhes neurofisiológicos são importantes, mas não devemos também adicionar um nível abstrato de descrição computacional que prescindia desses mesmos detalhes? Gallistel e King (2009) argumentam que uma fixação míope sobre o que nós atualmente sabemos sobre o cérebro tem conduzido a neurociência computacional a simplificar fenômenos cognitivos essenciais como a navegação, aprendizagem espaço-temporal e outros. Similarmente, Edelman (2014) reclama que o Enquadramento da Engenharia Neural substitui uma tempestade de detalhes neurofisiológicos por explicações psicológicas satisfatórias.

Parcialmente para responder tais preocupações, alguns pesquisadores propõem uma *neurociência cognitiva computacional* integrada que conecte teorias psicológicas com mecanismos de implementação neural (NASELARIS *et al.*, 2018; KRIEGESKORTE; DOUGLAS, 2018). A ideia básica é utilizar modelos de redes neurais para esclarecer como os processos mentais são instanciados no cérebro, desse modo, fundamentando a descrição cognitiva multiplamente realizável na descrição neurofisiológica. Um bom exemplo é o trabalho recente em implementação neural de inferências bayesianas [*Bayesian inference*] (por exemplo, POUGET *et al.*, 2013; ORHAN; MA, 2017; AITCHISON; LENGYEL, 2016). Pesquisadores articulam modelos bayesianos (multiplamente realizáveis) de vários processos mentais; eles constroem redes neurais biologicamente plausíveis que executam, ou executam aproximadamente, as computações bayesianas postuladas; e eles avaliam o quão bem esses modelos de redes neurais se adequam aos dados neurofisiológicos.

Apesar das diferenças entre o conexionismo e a neurociência computacional, esses dois movimentos fazem surgir muitos problemas parecidos. Em particular, a dialética da subseção 4.4 acerca da sistematicidade e da produtividade surgem de forma parecida.

## 5. Computação e representação

Filósofos e cientistas cognitivos usam o termo “representação” de modos diferentes. Dentro da filosofia, o uso mais dominante liga a representação à intencionalidade, ou seja, a **sobredade** [*aboutness*]<sup>14</sup> dos estados mentais. Filósofos contemporâneos geralmente explicam a intencionalidade invocando *conteúdo representacional*. Um estado mental representacional tem um conteúdo que representa o mundo como sendo de um certo modo, então podemos perguntar se o mundo é de fato desse modo. Assim, estados mentais com conteúdo em sentido representacional são *semanticamente avaliáveis* com respeito a propriedades como verdade, precisão [*accuracy*], cumprimento [*fulfillment*] e assim por diante. Para ilustrar:

- Crenças são o tipo de coisas que podem ser verdadeiras ou falsas. Minha crença de **que Emmanuel Macron é francês** é verdadeira se Emmanuel Macron for francês, falsa se ele não for.
- Estados perceptuais são o tipo de coisas que podem ser precisas ou imprecisas. Minha experiência perceptual **como a de uma esfera vermelha** é precisa apenas se uma esfera vermelha estiver diante de mim.
- Desejos são o tipo de coisas que podem ser cumpridas ou frustradas. Meu desejo de **comer chocolate** é cumprido se eu comer chocolate, frustrado se eu não comer chocolate.

Crenças têm condições de verdade (condições sob as quais elas são verdadeiras), estados perceptuais têm condições de precisão (condições sob as quais eles são precisos) e desejos têm condições de cumprimento (condições sob as quais eles são cumpridos).

---

<sup>14</sup> N.T.: Trata-se da capacidade que alguma coisa tem de ser sobre alguma outra coisa. A tradução mais comum do neologismo em inglês *aboutness* para o português é “acerca de”.

Na vida ordinária, nós frequentemente predizemos e explicamos comportamentos invocando crenças, desejos e outros tipos de estados mentais com conteúdo representacional. Nós identificamos esses estados por meio de suas propriedades representacionais. Quando dizemos “Frank acredita que Emmanuel Macron é francês”, nós especificamos as condições sob as quais a crença de Frank é verdadeira (a saber, que Emmanuel Macron seja francês). Quando nós dizemos “Frank quer comer chocolates”, nós especificamos as condições sob as quais o desejo de Frank é cumprido (a saber, que Frank coma chocolate). Então, a psicologia popular [*folk psychology*] atribui um papel central às *descrições intencionais*, ou seja, descrições que identificam estados mentais por meio de suas propriedades representacionais. Se a psicologia científica também deve empregar descrições intencionais é uma questão disputada dentro da filosofia da mente contemporânea.

O *realismo intencional* é o realismo com respeito à representação. No mínimo, essa posição sustenta que as propriedades representacionais são aspectos genuínos da mentalidade. Geralmente, também defende que a psicologia científica deveria empregar livremente as descrições intencionais quando apropriado. O realismo intencional é uma posição popular, defendida por Tyler Burge (2010a), Jerry Fodor (1987), Christopher Peacocke (1992; 1994) e muitos outros. Um argumento proeminente a favor do realismo intencional menciona as *práticas das ciências cognitivas*. O argumento sustenta que a descrição intencional figura centralmente em muitas áreas essenciais das ciências cognitivas, tais como a psicologia da percepção e a linguística. Por exemplo, a psicologia da percepção descreve como a atividade perceptual transforma as entradas sensoriais (por exemplo, estímulos retiniais) em representações de um ambiente distal (por exemplo, representações perceptuais de formatos, tamanhos e cores distais). A ciência identifica estados perceptuais por meio da menção a propriedades representacionais (por exemplo, relações representacionais a formatos, tamanhos e cores distais). Assumindo uma perspectiva realista científica ampla, as conquistas explicativas da psicologia da percepção dão suporte a uma postura realista com respeito à intencionalidade.

O *eliminativismo* é uma forma forte de antirrealismo sobre a intencionalidade. Eliminativistas atacam a descrição intencional qualificando-a como vaga, sensível ao contexto, relativa aos interesses, explicativamente superficial ou problemática por outros motivos. Eles recomendam que a psicologia científica descarte o conteúdo representacional. Um exemplo antigo é o *Word and Object* (1960), de W. V. Quine<sup>15</sup>, o qual busca substituir

---

<sup>15</sup> N.T.: *Palavras e Objeto*.

a psicologia intencional por uma psicologia behaviorista de estímulo-resposta. Paul Churchland (1981), outro eliminativista proeminente, quer substituir a psicologia intencional pela neurociência.

Entre o realismo intencional e o eliminativismo existem muitas posições intermediárias. Daniel Dennett (1971; 1987) reconhece que o discurso intencional é útil para predições, mas questiona se estados mentais *realmente* têm propriedades representacionais. Segundo Dennett, teóricos que empregam descrições intencionais não estão afirmando, **literalmente**, que estados mentais têm propriedades representacionais. Eles estão meramente adotando a “postura intencional”. Donald Davidson (1980) defende uma interpretação *interpretativista* (*interpretivist*) vizinha. Ele enfatiza o papel central que a atribuição intencional desempenha na prática interpretativa ordinária, ou seja, nossa prática de interpretar estados mentais e atos de fala uns dos outros. Ao mesmo tempo, ele questiona se a psicologia intencional irá encontrar um lugar dentro da teorização científica madura. Davidson e Dennett ambos professam o realismo sobre os estados mentais intencionais. Não obstante, ambos os filósofos são costumeiramente lidos como antirrealistas intencionais. Em particular, Dennett é frequentemente lido como um tipo de *instrumentalista* sobre a intencionalidade. Uma fonte dessa leitura habitual envolve a *indeterminação da interpretação*. Suponha que a evidência comportamental permita duas interpretações conflitantes dos estados mentais de um pensante. Seguindo Quine, Davidson e Dennett ambos dizem que não há então “nenhum fato relevante” acerca de qual interpretação está correta. Esse diagnóstico indica uma atitude não totalmente realista com respeito à intencionalidade.

Debates sobre intencionalidade figuram em destaque da discussão filosófica da TCM. Examinemos alguns pontos relevantes.

## 5.1 Computação como formal

Computacionalistas clássicos tipicamente assumem o que se pode chamar de *concepção sintático-formal da computação* (CSF). A ideia intuitiva é a de que a computação manipula símbolos em virtude de suas propriedades sintáticas formais e não de suas propriedades semânticas.

A CSF emerge das inovações na lógica matemática durante o fim do século XIX e início do XX, especialmente com as contribuições seminais de George Boole



e Gottlob Frege. Em sua *Begriffsschrift*<sup>16</sup> (1967[1879]), Frege efetuou uma **formalização** completa do raciocínio dedutivo. Para formalizar, especificamos uma **linguagem formal** cujas expressões linguísticas componentes são individuadas não-semanticamente (por exemplo, por seus formatos geométricos). Podemos ter em mente alguma interpretação pretendida, mas os elementos da linguagem formal são entidades puramente sintáticas que podemos discutir sem invocar propriedades semânticas, tais como referência ou condições de verdade. Em particular, podemos especificar **regras de inferência** em termos sintático-formais. Se escolhermos nossas regras de inferência com sabedoria, então elas serão coerentes com nossa interpretação pretendida: elas irão conduzir premissas verdadeiras a conclusões verdadeiras. Por meio de formalização, Frege investiu a lógica com um rigor sem precedentes. Com isso, ele estabeleceu o trabalho de base para inúmeros desenvolvimentos matemáticos e filosóficos subsequentes.

A formalização desempenha um papel fundacional significativo dentro da ciência da computação. Podemos programar um computador ao estilo Turing que manipula expressões linguísticas retiradas de uma linguagem formal. Se programarmos o computador com sabedoria, então suas maquinações sintáticas serão coerentes com nossa interpretação semântica pretendida. Por exemplo, podemos programar o computador de modo que ele conduza premissas verdadeiras apenas a conclusões verdadeiras, ou de modo que atualize probabilidades de acordo com o que preconiza a teoria da decisão bayesiana.

A CSF sustenta que **toda** a computação manipula itens sintático-formais, sem levar em conta quaisquer propriedades semânticas que esses itens possam ter. Formulações precisas da CSF variam. A computação é considerada como sendo “sensível” à sintaxe, mas não à semântica, ou como tendo “acesso” apenas a propriedades sintáticas, ou como operando “em virtude” de propriedades sintáticas ao invés das semânticas, ou como sendo impactada por propriedades semânticas apenas quando “mediadas” por propriedades sintáticas. Não é sempre claro o que essas formulações significam ou se são equivalentes entre si. Mas o quadro intuitivo é o de que propriedades sintáticas têm primazia causal/explicativa sobre as propriedades semânticas na condução da computação.

---

<sup>16</sup> N.T.: *Conceitografia*.

O artigo “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology” (1980) de Fodor<sup>17</sup> oferece uma formulação inicial. Fodor combina a CSF com a TCCM+TRM. Ele faz a analogia entre o mentalês e as linguagens formais estudadas pelos lógicos: o mentalês contém itens simples e complexos individuados não-semânticamente, assim como as linguagens formais tipicamente contém expressões simples e complexas individuadas por seus formatos. Símbolos em mentalês têm uma interpretação semântica, mas essa interpretação não impacta (diretamente) a computação mental. As propriedades formais de um símbolo, ao invés de suas propriedades semânticas, determinam como a computação manipula o símbolo. Nesse sentido, a mente é um “motor sintático”. Virtualmente todos os computacionalistas clássicos seguem Fodor no endosso da CSF.

Os conexionistas geralmente negam que redes neurais manipulam itens estruturados sintaticamente. Por essa razão, muitos conexionistas hesitariam em aceitar a CSF. Não obstante, muitos conexionistas endossam uma *tese de formalidade generalizada*: a computação é insensível a propriedades semânticas. A tese de formalidade generalizada levanta muitos dos mesmos problemas filosóficos levantados pela CSF. Focaremos aqui na CSF, a qual tem recebido mais atenção no debate filosófico.

Fodor combina a TCCM+TRM+CSF com o realismo intencional. Ele defende que a TCCM+TRM+CSF faz jus à psicologia popular ao ajudar a converter o discurso intencional do senso comum em uma ciência rigorosa. Ele justifica sua posição com o famoso argumento abduativo a favor da TCCM+TRM+CSF (1987, p. 18-20). Surpreendentemente, a atividade mental rastreia propriedades semânticas de forma coerente. Por exemplo, inferências dedutivas conduzem premissas a conclusões que são verdadeiras se as premissas são verdadeiras. Como podemos explicar esse aspecto crucial da atividade mental? A formalização mostra que as manipulações sintáticas são capazes de rastrear as propriedades semânticas, e a ciência da computação mostra como construir máquinas físicas que executam as manipulações sintáticas desejadas. Se tratarmos a mente como uma máquina orientada pela sintaxe, podemos explicar por que a atividade mental rastreia propriedades semânticas de forma coerente. Além do mais, nossa explicação não postula mecanismos causais radicalmente diferentes daqueles postulados dentro das ciências físicas. Respondemos, conseqüentemente, à questão crucial: **como a racionalidade é mecanicamente possível?**

---

<sup>17</sup> N.T.: “Solipsismo Metodológico considerado como uma Estratégia de Pesquisa em Psicologia Cognitiva”.

Stephen Stich (1983) e Hartry Field (2001) combinam a TCCM+CSF com o eliminativismo. Eles recomendam que as ciências cognitivas modelem a mente em termos sintático-formais, evitando totalmente a intencionalidade. Eles admitem que os estados mentais têm propriedades representacionais, mas se perguntam qual valor explicativo a psicologia científica ganha ao evocar essas propriedades. Por que suplementar a descrição sintático-formal com descrição intencional? Se a mente é uma máquina conduzida por sintaxe, então o conteúdo representacional não é dispensável por ser explicativamente irrelevante?

Em um ponto de sua carreira, Putnam (1983, p. 139-154) combinou a TCCM+CSF com um *interpretativismo* à la Davidson. Nessa concepção, as ciências cognitivas deveriam seguir as linhas sugeridas por Stich e Field, delineando modelos computacionais puramente sintático-formais. A modelagem sintático-formal coexiste com a prática interpretativa ordinária, na qual atribuímos conteúdos intencionais aos estados mentais e atos de fala de alguém. A prática interpretativa é governada por restrições holísticas e heurísticas, as quais impedem as tentativas de converter o discurso intencional em ciência rigorosa. Para Putnam, assim como para Field e Stich, a atividade científica ocorre no nível sintático-formal ao invés de no nível intencional.

A TCCM+CSF está sob ataque de várias direções. Uma linha de crítica mira na *relevância causal do conteúdo representacional* (BLOCK, 1990; FIGDOR, 2009; KAZEZ, 1995). Intuitivamente falando, os conteúdos dos estados mentais são causalmente relevantes para a atividade mental e para o comportamento. Por exemplo, meu desejo de beber água ao invés de suco de laranja causa que eu caminhe até à torneira e não até à geladeira. O conteúdo do meu desejo (**de beber água**) parece desempenhar um papel causal relevante na formação de meu comportamento. De acordo com Fodor (1990, p. 137-159), a TCCM+TRM+CSF acomoda tais intuições. A atividade sintático-formal **implementa** a atividade mental intencional, garantindo, por consequência, que os estados mentais intencionais interajam causalmente de acordo com seus conteúdos. No entanto, não fica claro se essa análise garante a relevância causal do conteúdo. A CSF diz que a computação é “sensível” à sintaxe, mas não à semântica. Dependendo de como alguém utiliza o termo chave “sensível”, pode parecer que o conteúdo representacional é irrelevante, com a sintaxe formal desempenhando todo o papel causal. Aqui vai uma analogia para ilustrar a preocupação. Quando um carro anda ao longo da estrada, há padrões estáveis envolvendo as sobras do carro. No entanto, a posição da sombra em um dado momento não influencia a posição da sombra de um momento posterior. De forma similar, a TCCM+TRM+CSF pode ser capaz de explicar como a atividade

mental instancia padrões estáveis descritos em termos intencionais, mas isso não é suficiente para garantir a relevância causal do conteúdo. Se a mente é uma máquina guiada pela sintaxe, então a eficácia causal parece residir no nível sintático e não no semântico. A semântica está apenas “pegando carona”. Aparentemente, portanto, a TCM+CSF encoraja a conclusão de que as propriedades representacionais são causalmente inertes. A conclusão pode não incomodar os eliminativistas, mas os realistas intencionais geralmente querem evitá-la.

Uma segunda linha de crítica recusa a imagem sintático-formal por considerá-la uma especulação sem fundamentos na prática científica. Tyler Burge (2010a; 2010b; 2013, p. 479-480) afirma que a descrição sintático-formal da atividade mental não desempenha papel relevante em grandes áreas das ciências cognitivas, incluindo o estudo do raciocínio teórico, do raciocínio prático e da percepção. Em cada caso, Burge argumenta, a ciência emprega a descrição intencional **ao invés** da descrição sintático-formal. Por exemplo, a psicologia da percepção individua estados perceptuais não por meio de propriedades sintático-formais, mas por meio de relações representacionais com formatos distais, tamanhos distais, cores distais e assim por diante. Para entender essa crítica, precisamos distinguir *descrição sintático-formal* e *descrição neurofisiológica*. Todos concordam que uma psicologia científica completa irá atribuir uma importância primordial à descrição neurofisiológica. No entanto, a descrição neurofisiológica é distinta da descrição sintático-formal, pois a descrição sintático-formal deve ser multiplamente realizável no plano neurofisiológico. O problema aqui é se a psicologia científica deveria complementar as *descrições intencionais* e as *descrições neurofisiológicas* com descrições *sintático-formais não intencionais multiplamente realizáveis*.

## 5.2 Externalismo<sup>18</sup> acerca do conteúdo mental

O artigo divisor de águas *The meaning of ‘Meaning’* (1975, p. 215-271) de Putnam<sup>19</sup> introduziu o *experimento de pensamento da Terra Gêmea*, o qual postula

---

<sup>18</sup>N.T.: O termo original em inglês *externalism* também é comumente traduzido como “externismo” na literatura filosófica lusófona. Sua contraparte, *internalism*, é traduzido por vezes como “internalismo”, por vezes como “internismo”.

<sup>19</sup>N.T.: O *Significado de ‘Significado’*.

um mundo exatamente com o nosso exceto que H<sub>2</sub>O é substituída por uma substância qualitativamente similar XYZ com uma composição química diferente. Putnam argumenta que XYZ não é água e que os falantes na Terra Gêmea usam a palavra **água** para se referir a XYZ e não a água. Burge (1982) estende essa conclusão da *referência linguística* para o *conteúdo mental*. Ele argumenta que os Terráqueos Gêmeos instanciam estados mentais com conteúdos diferentes. Por exemplo, se Oscar na Terra pensa *que a água mata a sede*, então seu duplicado na Terra Gêmea pensa um pensamento com conteúdo diferente, que podemos colocar como **que a água mata a sede**. Burge concluiu que o conteúdo mental não sobrevém à neurofisiologia interna. O conteúdo mental é individuado parcialmente por fatores externos à pele do pensante, incluindo relações causais com o ambiente. Essa posição é o *externalismo sobre o conteúdo mental*.

Propriedades sintático-formais dos estados mentais são amplamente tidas como supervenientes à neurofisiologia interna. Por exemplo, Oscar e Oscar Gêmeo instanciam as mesmas manipulações sintático-formais. Assumindo o externalismo de conteúdo, segue-se que há um enorme abismo entre a descrição intencional ordinária e a descrição sintático-formal.

O externalismo de conteúdo levanta sérias questões sobre a utilidade explicativa do conteúdo representacional para a psicologia científica:

O *Argumento da Causalidade* [*Argument from Causation*] (FODOR, 1987; 1991): como um conteúdo mental pode exercer qualquer influência causal exceto conforme manifestado dentro da neurofisiologia interna? Não há “ação psicológica à distância”. Diferenças no ambiente físico impactam o comportamento apenas quando induzem diferenças em estados cerebrais locais. Então, os únicos fatores causalmente relevantes são aqueles que sobrevém à neurofisiologia interna. O conteúdo externamente individuado é **causalmente irrelevante**.

O *Argumento da Explicação* [*Argument from Explanation*] (STICH, 1983): explicações científicas rigorosas não deveriam levar em conta fatores externos à pele do sujeito. A psicologia popular pode fornecer uma taxonomia dos estados mentais por meio de relações com o ambiente externo, porém a psicologia científica deveria fornecer uma taxonomia dos estados mentais inteiramente a partir de fatores que sobrevém à neurofisiologia interna. Deveria tratar Oscar e Oscar Gêmeo como sendo psicologicamente duplicados<sup>20</sup>.

---

<sup>20</sup> N.T.: Um argumento semelhante alega apenas que a explicação internalista oferece algumas vantagens sobre a explicação externalista (BLOCK, 1986; CHALMERS, 2002; LEWIS, 1994). Esse argumento não tenta expurgar o conteúdo amplo da explicação psicológica. Simplesmente defende que ganhamos benefícios explicativos mencionando o conteúdo exíguo.

Alguns autores desenvolvem os dois argumentos em conjunto. Ambos levam à mesma conclusão: o conteúdo mental externamente individuado não encontra um lugar legítimo dentro da explicação causal fornecida pela psicologia científica. Stich (1983) argumenta nessa linha para fundamentar seu eliminativismo sintático-formal.

Muitos filósofos respondem a essas preocupações promovendo o *internalismo de conteúdo*. Se externalistas de conteúdo favorecem o *conteúdo amplo* (*wide content*) (o conteúdo que não sobrevém à neurofisiologia interna), internalistas de conteúdo favorecem o *conteúdo exíguo* (*narrow content*) (o conteúdo que sobrevém à neurofisiologia interna). O conteúdo exíguo é o que permanece do conteúdo mental quando se retira todos os elementos externos. Em um ponto de sua carreira, Fodor (1981; 1987) seguiu o internalismo como uma estratégia para integrar a psicologia intencional com a TCCM+TRM+CSF. Ao mesmo tempo que concedia que o conteúdo amplo não devesse figurar em uma psicologia científica, sustentava que o conteúdo exíguo deveria desempenhar um papel explicativo central.

Internalistas radicais insistem que *todo* o conteúdo é exíguo. Uma análise típica nessa linha sustenta que Oscar está pensando não sobre água, mas sobre algo de uma categoria de substância mais geral que englobe XYZ, assim, Oscar e Oscar Gêmeo entretêm estados mentais como os mesmos conteúdos. Tim Crane (1991) e Gabriel Segal (2000) endossam uma tal análise. Eles sustentam que a psicologia popular sempre individua as atitudes proposicionais do modo exíguo. Um internalismo menos radical recomenda que reconheçamos o conteúdo exíguo **em conjunto** com o conteúdo amplo. A psicologia popular pode, às vezes, individuar atitudes proposicionais do modo amplo, mas podemos também delinear uma noção viável de conteúdo exíguo que avança objetivos filosóficos ou científicos importantes. Internalistas têm proposto várias noções candidatas a conteúdo exíguo (BLOCK, 1986; CHALMERS, 2002; CUMMINS, 1989; FODOR, 1987; LEWIS, 1994; LOAR, 1988; MENDOLA, 2008). Veja o verbete da SEP, **narrow mental content**, para uma visão geral dos candidatos proeminentes.

Externalistas reclamam que as teorias existentes acerca do conteúdo exíguo são apenas rascunhos, são implausíveis, são inúteis para a explicação psicológica, ou, senão, são objetáveis (BURGE, 2007; SAWYER, 2000; STALNAKER, 1999). Externalistas também questionam os argumentos internalistas de que a psicologia científica requer conteúdo exíguo.

*Argumento da Causalidade:* Externalistas insistem que o conteúdo amplo pode ser causalmente relevante. Os detalhes variam entre os externalistas e a

discussão frequentemente se torna entrelaçada a problemas complexos que cercam a causalidade, os contrafactuais e a metafísica da mente. Veja o verbete da SEP, **mental causation**, para uma visão introdutória e veja Burge (2007), Rescorla (2014a) e Yablo (1997, 2003) para uma discussão externalista representativa.

*Argumento da Explicação:* Externalistas alegam que a explicação psicológica pode legitimamente fornecer uma taxonomia dos estados mentais por meio de fatores que ultrapassam a neurofisiologia interna (PEACOCKE, 1993; SHEA, 2018). Burge pontua que as ciências não psicológicas geralmente individualizam espécies explicativas **de modo relacional**, ou seja, por meio de relações com fatores externos. Por exemplo, se uma entidade conta como um coração depende (grosso modo) de se sua função biológica em seu ambiente normal seja bombear sangue. Então a fisiologia individualiza espécies de órgãos relacionalmente. Por que a psicologia não pode, do mesmo modo, individualizar os estados mentais relacionalmente? Para um debate notável dessas questões, veja Burge (1986; 1989; 1995) e Fodor (1987; 1991).

Externalistas duvidam que tenhamos qualquer boa razão para substituir ou suplementar o conteúdo amplo com o conteúdo exíguo. Eles rejeitam a busca pelo conteúdo exíguo por considerá-la uma busca infrutífera.

Burge (2007; 2010a) defende o externalismo analisando a ciência cognitiva atual. Ele argumenta que muitos ramos da psicologia científica (especialmente a psicologia da percepção) individualizam o conteúdo mental por meio das relações causais com o ambiente externo. Ele conclui que a prática científica incorpora uma perspectiva externalista. Por outro lado, ele mantém que o conteúdo exíguo é uma fantasia filosófica sem fundamento na ciência atual.

Suponhamos que abandonemos a busca por um conteúdo exíguo. Quais são os candidatos para combinar a TSM+CSF com a psicologia intencional externalista? A opção mais promissora enfatiza **níveis de explicação**. Podemos dizer que a psicologia intencional ocupa um nível de explicação, enquanto a psicologia computacional sintático-formal ocupa um nível diferente. Fodor defende essa abordagem em seus trabalhos posteriores (1994; 2008). Ele passou a rejeitar o conteúdo exíguo por considerá-lo ocioso. Ele sugere que os mecanismos sintático-formais implementam leis psicológicas externalistas. A computação mental manipula expressões em mentais de acordo com suas propriedades sintático-formais e essas manipulações sintático-formais garantem que a atividade mental instancie os padrões regulares apropriados definidos por meio dos conteúdos amplos.

À luz da distinção internalismo/externalismo, revisitemos o desafio eliminativista levantado na subseção 5.1: qual valor explicativo a descrição intencional

adiciona à descrição sintático-formal? Internalistas podem responder que manipulações sintático-formais adequadas determinam, e até mesmo constituem, conteúdos exíguos, de modo que a descrição intencional internalista já está implícita na descrição sintático-formal (*vide* FIELD, 2001, p. 75). Talvez essa resposta justifique o realismo intencional, talvez não. O ponto crucial, no entanto, é que tal resposta não está disponível aos externalistas de conteúdo. A descrição intencional externalista não está implícita na descrição sintático-formal, porque se pode manter fixa a sintaxe formal enquanto se varia o conteúdo amplo. Portanto, externalistas de conteúdo que simpatizam com a TCM+CSF precisam dizer o que ganhamos ao suplementar as explicações sintático-formais com explicações intencionais. Uma vez que se aceite que a computação mental é sensível à sintaxe, mas não à semântica, está muito longe de claro que reste ao conteúdo amplo qualquer trabalho explicativo útil. Fodor aborda esse desafio em vários momentos, oferecendo seu mais sistemático tratamento em *The Elm and the Expert* (1994)<sup>21</sup>. Veja Arjo (1996), Aydede (1998), Aydede e Robbins (2001), Wakefield (2002), Perry (1998) e Wakefield (2002) para críticas. Veja Rupert (2008) e Schneider (2005) para posições próximas à de Fodor. Dretske (1993) e Shea (2018, p. 197-226) percorrem estratégias alternativas para reivindicar relevância explicativa ao conteúdo amplo.

### 5.3 Computação envolvedora-de-conteúdo

O abismo perceptível entre a descrição computacional e a descrição intencional estimula muitos escritos em TCM. Alguns filósofos tentam construir a ponte sob o abismo usando descrições computacionais que individualizam estados computacionais em termos representacionais. Essas descrições são *envolvedoras-de-conteúdo* [*content-involving*], para usar a terminologia de Christopher Peacocke (1994). Na abordagem envolvedora-de-conteúdo, não há uma demarcação rígida entre a descrição computacional e a intencional. Em particular, certas descrições da atividade mental cientificamente valiosas são tanto computacionais quanto intencionais. Chame essa posição de *computacionalismo envolvedor-de-conteúdo*.

---

<sup>21</sup> N.T.: O Ulmeiro e o Especialista.



O computacionalismo envolvedor-de-conteúdo não precisa dizer que toda a descrição computacional é intencional. Para ilustrar, suponha que descrevamos uma máquina de Turing simples que manipule símbolos individuados por seus formatos geométricos. Então, a descrição computacional resultante plausivelmente não será envolvedora-de-conteúdo. Portanto, computacionalistas envolvidos-de-conteúdo geralmente não desenvolvem a computação envolvedora-de-conteúdo como uma teoria geral da computação. Eles alegam apenas que **algumas** descrições computacionais importantes são envolvedoras-de-conteúdo.

Pode-se desenvolver um computacionalismo envolvedor-de-conteúdo em uma direção internalista ou externalista. *Computacionalistas envolvidos-de-conteúdo internalistas* sustentam que algumas descrições computacionais identificam os estados mentais parcialmente por meio de seus conteúdos *exíguos*. Murat Aydede (2005) recomenda uma posição nessa linha. *Computacionalistas envolvidos-de-conteúdo externalistas* sustentam que certas descrições computacionais identificam os estados mentais parcialmente por meio de seus conteúdos *amplos*. Tyler Burge (2010a, p. 95-101), Christopher Peacocke (1994; 1999), Michael Rescorla (2012) e Mark Sprevak (2010) simpatizam com essa posição. Oron Shagrir (2001; 2020) defende um computacionalismo envolvedor-de-conteúdo que é neutro com relação ao internalismo e ao externalismo.

Computacionalistas envolvidos-de-conteúdo externalistas tipicamente citam a prática das ciências cognitivas como um fator motivador. Por exemplo, a psicologia da percepção descreve o sistema perceptual como computando uma estimativa do tamanho de algum objeto a partir dos estímulos retiniais e de uma estimativa da profundidade do objeto. As “estimativas” perceptuais são identificadas representacionalmente, como representações de tamanhos e profundidades distais específicos. Muito plausivelmente, relações representacionais com tamanhos e profundidades distais não sobrevivem à neurofisiologia interna. Muito plausivelmente, então, a psicologia da percepção identifica as computações perceptuais em tipos por meio de seu conteúdo amplo. Portanto, o computacionalismo envolvedor-de-conteúdo externalista parece harmonizar-se bem com as ciências cognitivas atuais.

Um grande desafio enfrentado pelo computacionalismo envolvedor-de-conteúdo diz respeito à interface com os formalismos padrão do computacionalismo, tais como a máquina de Turing. Como exatamente as descrições envolvedoras-de-conteúdo se relacionam com os modelos computacionais encontrados na lógica e na ciência da computação? Filósofos geralmente assumem que esses modelos oferecem descrições não intencionais. Se for assim, esse seria um grande, e talvez

fatal, golpe no computacionalismo envolvente-de-conteúdo.

É possível argumentar, no entanto, que muitos formalismos computacionais familiares permitem uma construção envolvente-de-conteúdo ao invés de uma construção sintático-formal. Para ilustrar, considere a máquina de Turing. **Pode-se** individualizar não semanticamente os “símbolos” que compõem o alfabeto da máquina de Turing, por meio de fatores semelhantes a um formato geométrico. Mas o formalismo de Turing **requer** um esquema de individualização não semântico? Pode-se argumentar que o formalismo permite que individualizemos símbolos parcialmente por meio de seus conteúdos. Obviamente, a tabela de máquina para uma máquina de Turing não menciona explicitamente propriedades semânticas de símbolos (por exemplo, denotações ou condições de verdade). Não obstante, a tabela de máquina pode codificar regras mecânicas que descrevem como manipular símbolos, onde esses símbolos sejam identificados quanto ao tipo em termos que envolvam conteúdo. Dessa forma, a tabela de máquina dita transições entre os estados envolventes-de-conteúdo sem mencionar explicitamente propriedades semânticas. Aydede (2005) sugere uma versão internalista dessa visão, com símbolos identificados quanto ao tipo por meio de seu conteúdo exíguo<sup>22</sup>. Rescorla (2017a) desenvolve uma visão em uma direção externalista, com símbolos identificados quanto ao tipo por meio de seus conteúdos amplos. Ele argumenta que alguns modelos ao estilo Turing descrevem operações computacionais sobre símbolos em mentais individualizados de modo externalista<sup>23</sup>.

---

<sup>22</sup> A discussão inicial de Fodor (1980) sugeria uma visão similar: dois espécimes de um único tipo sintático em mentais compartilhariam o mesmo conteúdo exíguo, mas não necessariamente o mesmo conteúdo amplo. Por exemplo, haveria um tipo sintático em mentais ÁGUA que poderia denotar tanto H<sub>2</sub>O quanto XYZ, mas que necessariamente expressaria um único conteúdo exíguo fixo. A computação mental é formal (pois é insensível a propriedades semânticas determinadas externamente) e envolventes-de-conteúdo-exíguo (pois os tipos sintáticos em mentais têm em essência seus conteúdos exíguos). O trabalho posterior de Fodor (do início da década de 1990 em diante) abandona o conteúdo exíguo e, conjuntamente, abandona toda inclinação em favor de uma computação envolvente-de-conteúdo.

<sup>23</sup> Horowitz (2007), Bontly (1998) e Shea (2013) também favorecem a individualização externalista dos veículos computacionais, embora por razões um pouco diferentes das consideradas aqui.

À princípio, pode-se abraçar em conjunto a descrição computacional envolvente-de-conteúdo externalista e a descrição sintático-formal. Pode-se dizer que esses dois tipos de descrição ocupam níveis distintos de explicação. Peacocke sugere uma visão assim. Outros computacionalistas envolvidos-de-conteúdo se mantêm mais céticos com respeito às descrições sintático-formais da mente. Por exemplo, Burge questiona com qual valor explicativo a descrição sintático-formal pode contribuir para certas áreas da psicologia científica (tal como a psicologia da percepção). Segundo esse ponto de vista, o desafio eliminativista posto na subseção 5.1 tem um efeito oposto. Não deveríamos assumir que as descrições sintático-formais são explicativamente valiosas e então perguntar com o que as descrições intencionais podem contribuir. Deveríamos, ao contrário, abraçar as descrições intencionais externalistas oferecidas pelas ciências cognitivas atuais e, então, perguntar com o que a descrição sintático-formal pode contribuir.

Proponentes da descrição sintático-formal respondem mencionando *mecanismos de implementação*. A descrição externalista da atividade mental pressupõe que relações histórico-causais adequadas entre a mente e o ambiente físico externo estão em jogo. Mas certamente queremos uma descrição “local” que ignore as relações histórico-causais externas, uma descrição que revele os mecanismos causais subjacentes. Fodor (1987; 1994) argumenta nessa linha para justificar o panorama sintático-formal. Para respostas externalistas possíveis ao argumento dos mecanismos de implementação, veja Burge (2010b), Rescorla (2017b), Shea (2013) e Sprevak (2010). Debates sobre esse argumento e, de um modo mais geral, sobre a relação entre computação e representação, parecem estar fadados a continuar por um futuro indefinido.

## 6. Concepções alternativas de computação

A literatura oferece diversas concepções alternativas, geralmente levadas adiante como fundamentos para a TCM. Em muitos casos, essas concepções se sobrepõem umas às outras ou às concepções consideradas anteriormente.

## 6.1 Processamento de informação

É comum que cientistas cognitivos descrevam computação como **processamento de informação**. É menos comum que os proponentes clarifiquem o que querem dizer com “processamento” e “informação”. Sem clarificação, a descrição é pouco mais que um *slogan* vazio.

Claude Shannon introduziu uma noção cientificamente importante de “informação” em seu artigo *A Mathematical Theory of Communication*<sup>24</sup>, de 1948. A ideia intuitiva é a de que a informação mede a *redução da incerteza* [*reduction in uncertainty*], em que a incerteza reduzida se manifesta como uma distribuição de probabilidade alterada sobre estados possíveis. Shannon codificou essa ideia em um rigoroso esquema matemático, lançando as bases da *teoria da informação* (COVER; THOMAS, 2006). A informação de Shannon é fundamental para a engenharia moderna. Ela encontra aplicação frutífera dentro das ciências cognitivas, especialmente das neurociências cognitivas. Ela apoia uma análise convincente da computação em termos de “processamento de informação”? Considere um toca fitas antiquado que grava mensagens recebidas por um rádio sem fio. Usando o esquema de Shannon, pode-se medir o quanto de informação é carregada por alguma mensagem gravada. Há um sentido em que o toca fitas “processa” a informação de Shannon quando reproduzimos uma mensagem gravada. Ainda assim, a máquina não parece implementar um modelo computacional não trivial<sup>25</sup>. Certamente, nem o formalismo da máquina de Turing nem o formalismo das redes neurais oferecem muitos esclarecimentos sobre as operações na máquina. Pode-se argumentar, então, que um sistema pode processar informações de Shannon sem executar computações em qualquer sentido interessante.

Confrontando-se com tais exemplos, pode-se tentar isolar uma noção mais robusta de **processamento**, de modo que o toca fitas não “processe” informações de Shannon. Alternativamente, pode-se insistir que o toca fitas executa computações não triviais. Piccinini e Scarantino (2010) desenvolvem uma noção altamente geral

---

<sup>24</sup> N.T.: *Uma Teoria Matemática da Comunicação*.

<sup>25</sup> Como discutido mais à frente na subseção 7.1, a máquina pode implementar um modelo computacional trivial (por exemplo, um autômato finito de um único estado). No entanto, um modelo trivial desse tipo ilumina muito pouco, se é que ilumina alguma coisa, a respeito das operações da máquina.

de computação que tem essa consequência, a qual chamam de *computação genérica*.

Uma segunda noção proeminente de informação deriva da influente discussão de Paul Grice (1989) sobre *significado natural*. Significado natural envolve correlações que suportem contrafactuais e sejam confiáveis. Por exemplo, os anéis dos troncos das árvores correlacionam-se com a idade da árvore e feridas de varicela se correlacionam com varicela. Coloquialmente descrevemos anéis de troncos de árvores como carregando informações sobre a idade da árvore, feridas de varicela como carregando informação sobre varicela e assim por diante. Tais descrições sugerem uma concepção que atrela a informação a correlações que suportam contrafactuais e são confiáveis. Fred Dretske (1981) desenvolve essa concepção em uma teoria sistemática, assim como o fazem vários filósofos posteriores. A informação ao estilo de Dretske é capaz de suprir uma análise plausível da computação como “processamento de informação”? Considere um antiquado *termostato bimetalico*. Dois metais são unidos em uma chapa. A dilatação diferenciada dos metais faz com que a chapa se curve, desse modo ativando ou desativando uma unidade de aquecimento. O estado da chapa se correlaciona confiavelmente com a temperatura atual do ambiente e o termostato “processa” esses estados que têm informação quando ativa ou desativa o aquecedor. Ainda assim, o termostato não parece implementar qualquer modelo computacional não trivial. Normalmente, não se consideraria que o termostato computa. Pode-se argumentar, então, que um sistema pode processar informações ao estilo Dretske sem executar computações em qualquer sentido interessante. Claro, pode-se tentar lidar com esses exemplos por meio de artimanhas paralelas às do parágrafo anterior.

Uma terceira noção proeminente de informação é a **informação semântica**, ou seja, conteúdo representacional<sup>26</sup>. Alguns filósofos sustentam que um sistema físico computa apenas se os estados desse sistema têm propriedades representacionais (DIETRICH, 1989; FODOR, 1998, p. 10; LADYMAN, 2009; SHAGRIR, 2006; SPREVAK, 2010). Nesse sentido, o processamento de informação é **necessário** para a computação. Como Fodor (1975, p. 34) memoravelmente coloca, “não há computação sem representação”. No entanto, essa posição é discutível. Chalmers (2011) e Piccinini (2008a) afirmam que uma máquina de Turing pode executar

---

<sup>26</sup> Filósofos de mentalidade naturalista geralmente tentam reduzir o conteúdo representacional à informação ao estilo Dretske (DRETSKE, 1981; FODOR, 1990). Esse projeto reducionista é controverso.

computações mesmo que os símbolos manipulados pela máquina não tenham interpretação semântica. As computações da máquina são puramente sintáticas em sua natureza, desprovidas de qualquer coisa próxima a propriedades semânticas. Nessa visão, o conteúdo representacional não é necessário para que um sistema físico conte como computacional.

Ainda não é claro se o *slogan* “computação é processamento de informação” fornece alguma revelação relevante. Não obstante, parece improvável que essa frase desapareça da literatura tão cedo. Para uma maior discussão sobre as possíveis conexões entre computação e informação, veja Gallistel e King (2009, p. 1-26), Lizier, Flecker e Williams (2013), Milkowski (2013), Piccinini e Scarantino (2010) e Sprevak (2020).

## 6.2 Resolução de funções

Em uma passagem amplamente citada, o psicólogo da percepção David Marr (1982) distingue três níveis por meio dos quais se pode descrever um “dispositivo de processamento de informação”:

*Teoria computacional*: “o dispositivo é caracterizado como mapeando de um tipo de informação para outro, as propriedades abstratas desse mapeamento são definidas com precisão e sua conveniência e adequação para a tarefa em questão são demonstradas” (p. 24).

*Representação e algoritmo*: “a escolha da representação para a entrada e para a saída e o algoritmo utilizado para transformar uma na outra” (p. 24-25).

*Implementação física [hardware implementation]*: “os detalhes de como o algoritmo e a representação são fisicamente realizados” (p. 25).

Os três níveis de Marr têm atraído intenso escrutínio filosófico. Para nossos propósitos, o ponto central é que o “nível computacional” de Marr descreve um mapeamento das entradas às saídas, sem descrever passos intermediários. Marr ilustra sua abordagem providenciando teorias do “nível computacional” de vários processos perceptuais, tais como a detecção de bordas.

A discussão de Marr sugere uma *concepção funcional da computação*, na qual a computação é uma questão de transformação de entradas em saídas apropriadas. Frances Egan elabora a concepção funcional em uma série de artigos (1991; 1992; 1999; 2003; 2010; 2014; 2019). Como Marr, ela trata a descrição computacional como a descrição das relações entre entradas e saídas. Ela também alega que modelos computacionais descrevem uma função puramente **matemática**: isso é, um mapeamento de entradas matemáticas para saídas matemáticas. Ela ilustra isso considerando um mecanismo visual (chamado *Visua*) que computa a profundidade de um objeto a partir da disparidade retinal. Ela imagina um duplicado neurofisiológico (*Visua Gêmeo*) situado de maneira tão diferente no ambiente físico que ele não representa a profundidade. *Visua* e *Visua Gêmeo* instanciam estados perceptuais com diferentes propriedades representacionais. Não obstante, Egan diz, a ciência da visão trata *Visua* e *Visua Gêmeo* como **duplicados computacionais**. *Visua* e *Visua Gêmeo* computam a mesma função matemática, embora as computações tenham teores representacionais diferentes nos dois casos. Egan conclui que a modelagem computacional da mente produz uma “descrição matemática abstrata” consistente com muitas descrições representacionais possíveis distintas. A atribuição intencional é apenas uma formulação heurística sobre uma descrição computacional subjacente.

Chalmers (2012) argumenta que a concepção funcional negligencia características importantes da computação. Como ele observa, modelos computacionais geralmente descrevem mais do que apenas relações entre entradas e saídas. Eles descrevem passos intermediários através dos quais as entradas são transformadas em saídas. Esses passos intermediários, os quais Marr aloca no nível **algorítmico**, figuram proeminentemente nos modelos computacionais oferecidos por lógicos e cientistas da computação. Restringir o termo **computação** a descrições de entrada e saída não captura a prática computacional padrão.

Uma preocupação adicional é enfrentada pelas teorias funcionais, tais como a de Egan, que enfatizam exclusivamente as entradas e saídas **matemáticas**. Críticos reclamam que Egan erroneamente superestima as funções matemáticas, às custas das explicações intencionais rotineiramente oferecidas pelas ciências cognitivas (BURGE, 2005; RESCORLA, 2015; SILVERBERG, 2006; SPREVAK, 2010). Para ilustrar, suponha que a psicologia da percepção descreva o sistema perceptual como estimando que a profundidade de algum objeto é 5 metros. A estimativa de profundidade perceptual tem um conteúdo representacional: é precisa apenas se a profundidade do objeto for 5 metros. Nós mencionamos o número 5 para identificar a estimativa de profundidade. Mas nossa escolha do número depende

de escolhas arbitrárias acerca das unidades de medidas. Críticos afirmam que o conteúdo da estimativa de profundidade – não o número escolhido arbitrariamente a partir do qual nós teóricos especificamos esse conteúdo – é o que importa para a explicação psicológica. A teoria de Egan põe o número, ao invés do conteúdo, no centro da explicação. De acordo com Egan, a explicação computacional deveria descrever o sistema visual como computando uma **função matemática específica** que conduz **entradas matemáticas específicas a saídas matemáticas específicas**. Essas entradas e saídas matemáticas específicas dependem das nossas escolhas arbitrárias acerca de unidades de medida, então, podemos argumentar, elas carecem da significância explicativa que Egan atribui a elas.

Devemos distinguir a abordagem funcional, como a percorrida por Marr e Egan, do *paradigma de programação funcional* [*functional programming paradigma*] da ciência da computação. O paradigma de programação funcional modela a resolução de uma função complexa em sucessivas resoluções de funções simples. Para pegar um exemplo simples, pode-se resolver  $f(x,y)=(x^2+y)$  resolvendo primeiramente a função quadrática e então resolvendo a função de adição. A programação funcional difere das descrições de “nível computacional” enfatizadas por Marr, pois ela especifica etapas computacionais intermediárias. O paradigma da programação funcional remonta ao *Cálculo Lambda*, de Alonzo Church (1936), continuado com linguagens de programação como PCF e LISP. Ele desempenha um papel importante na IA e na ciência da computação teórica. Alguns autores sugerem que ele seja capaz de iluminar a investigação sobre a computação mental (KLEIN, 2012; PIANTADOSI; TENENBAUM; GOODMAN, 2012). No entanto, muitos formalismos computacionais não estão em conformidade com o paradigma funcional: máquinas de Turing; linguagens de programação imperativas, como a C; linguagens de programação lógicas, como a Prolog; e assim por diante. Embora o paradigma funcional descreva numerosas computações importantes (incluindo possivelmente computações mentais), ela não captura plausivelmente a computação **em geral**.

### 6.3 Estruturalismo

Muitas discussões filosóficas incorporam uma *concepção estruturalista da computação*: um modelo computacional descreve uma estrutura causal abstrata, sem levar em conta estados físicos específicos que instanciam a estrutura. Essa concepção remonta pelo menos até ao tratamento original de Putnam (1967).



Chalmers (1995; 1996a; 2011; 2012) a desenvolve em detalhes. Ele introduz o formalismo de *autômato de estado combinatório* (AEC) [*combinatorial-state automaton* (CSA)] –, que engloba os modelos mais familiares de computação (incluindo as máquinas de Turing e as redes neurais). Um AEC provê uma descrição abstrata da *topologia causal* de um sistema físico: o padrão de interação causal entre as partes do sistema, independente da natureza dessas partes ou do mecanismo causal por meio do qual elas interagem. A descrição computacional especifica a topologia causal.

Chalmers emprega o estruturalismo para delinear uma versão muito geral da TCM. Ele assume a visão funcionalista segundo a qual os estados mentais são individuados por meio de seus papéis num padrão de organização causal. A descrição psicológica especifica papéis causais, abstraídos dos estados físicos que realizam esses papéis. Assim, propriedades psicológicas são **organizacionalmente invariantes**, no sentido de que elas sobrevivem à topologia causal. Visto que a descrição computacional caracteriza uma topologia causal, satisfazer uma descrição computacional adequada é suficiente para instanciar propriedades mentais apropriadas. Segue-se também que a descrição psicológica é uma espécie da descrição computacional, de tal modo que a descrição computacional deveria desempenhar um papel central dentro da explicação psicológica. Portanto, a computação estruturalista provê uma base sólida para as ciências cognitivas. A mentalidade é fundada em padrões causais, os quais são precisamente o que os modelos computacionais articulam.

O estruturalismo vem embalado com uma explicação atraente da **relação de implementação** entre modelos computacionais e sistemas físicos. Sob quais condições um sistema físico implementa um modelo computacional? Estruturalistas dizem que um sistema físico implementa um modelo apenas no caso de a estrutura causal do modelo ser **isomórfica** à estrutura formal do modelo. Um modelo computacional descreve um sistema físico articulando uma estrutura formal que espelha alguma topologia causal relevante. Chalmers desenvolve essa ideia intuitiva, provendo detalhadas condições necessárias e suficientes para a realização dos AECs. Poucas, se é que alguma, concepções alternativas de computação podem fornecer uma explicação tão substancial acerca da relação de implementação.

Podemos comparar instrutivamente o computacionalismo estruturalista com algumas outras teorias discutidas anteriormente:

*Funcionalismo de máquina.* O computacionalismo estruturalista abraça a ideia central por trás do funcionalismo de máquina: estados mentais são estados funcionais descritíveis por meio de um formalismo computacional adequado. Putnam

avança a TCM como uma hipótese empírica e defende o funcionalismo com base nisso. Ao contrário, Chalmers segue David Lewis (1972) ao fundamentar o funcionalismo na análise conceitual do discurso mentalista. Enquanto Putnam defende o funcionalismo por meio de uma defesa do computacionalismo, Chalmers defende o computacionalismo assumindo o funcionalismo.

*Computacionalismo clássico, conexionismo e a neurociência computacional.* O computacionalismo estruturalista enfatiza as descrições organizacionalmente invariantes, as quais são multiplamente realizáveis. Com respeito a esse ponto, diverge das neurociências computacionais. O estruturalismo é compatível tanto com o computacionalismo clássico quanto com o conexionista, mas difere em espírito dessas visões. Classicistas e conexionistas apresentam suas posições rivais como hipóteses substanciais e ousadas. Chalmers elabora o computacionalismo estruturalista como uma posição relativamente minimalista difícil de ser refutada.

*Realismo intencional e eliminativismo.* O computacionalismo estruturalista é compatível com ambas as posições. A descrição do AEC não menciona explicitamente propriedades semânticas tais como referência, condições de verdade, conteúdo representacional e assim por diante. Computacionalistas estruturalistas não precisam atribuir ao conteúdo representacional qualquer papel importante dentro da psicologia científica. Por outro lado, o computacionalismo estruturalista também não impede que seja atribuído um papel relevante ao conteúdo representacional.

*A concepção sintático-formal da computação.* O conteúdo amplo depende de relações histórico-causais com o ambiente externo, relações que ultrapassam a topologia causal. Assim, a descrição do AEC deixa o conteúdo amplo indeterminado. Presumivelmente, o conteúdo exíguo sobrevém à topologia causal, mas a descrição do AEC não menciona explicitamente conteúdos exíguos. De um modo geral, portanto, o computacionalismo estruturalista prioriza um nível de descrição computacional não semântica e formal. Nesse sentido, se parece com a CSF. Por outro lado, computacionalistas estruturalistas não precisam dizer que a computação é “insensível” às propriedades semânticas, portanto, eles não precisam endossar todos os aspectos da CSF.

Embora o computacionalismo estruturalista seja distinto da TCM+CSF, ele levanta alguns problemas similares. Por exemplo, Rescorla (2012) nega que a topologia causal desempenhe o papel explicativo central dentro das ciências cognitivas ditado pelo computacionalismo estruturalista. Ele sugere que é a descrição intencional externalista, ao invés da descrição organizacionalmente invariante, que desfruta de primazia explicativa. Vindo de uma direção diferente, neurocientistas

computacionais recomendam que renunciemos às descrições organizacionalmente invariantes e, no lugar, empreguemos modelos computacionais mais especificamente neuronais. Em resposta a essas objeções, Chalmers (2012) argumenta que a descrição computacional organizacionalmente invariante rende benefícios explicativos que nem a descrição intencional nem a descrição neurofisiológica replicam: ela revela os mecanismos subjacentes à cognição (diferentemente da descrição intencional) e ela se abstrai dos detalhes da implementação neural que são irrelevantes para muitos fins explicativos.

#### 6.4 Teorias mecanicistas

A natureza mecanicista da computação é um tema recorrente em lógica, filosofia e ciências cognitivas. Gualtiero Piccinini (2007; 2012; 2015) e Marcin Milkowski (2013) evoluíram esse tema para uma teoria mecanicista dos sistemas computacionais. Um *mecanismo funcional* é um sistema de componentes interconectados, onde cada componente desempenha alguma função dentro do sistema geral. A *explicação mecanicista* procede decompondo o sistema em partes, descrevendo como as partes são organizadas dentro do sistema maior e isolando a função desempenhada por cada parte. Um sistema computacional é um mecanismo funcional de um tipo específico. Na abordagem de Piccinini, um sistema computacional é um mecanismo cujos componentes são organizados funcionalmente para processar veículos de acordo com regras. Ecoando a discussão de Putnam sobre a múltipla realizabilidade, Piccinini exige que as regras sejam **independentes do meio**, no sentido de que elas sejam abstratas em relação à implementação física específica dos veículos. A explicação computacional decompõe o sistema em partes e descreve como cada parte ajuda o sistema a processar os veículos relevantes. Se o sistema processa veículos cuja estrutura é discreta, então a computação é digital. Se o sistema processa veículos contínuos, então a computação é analógica. A versão da abordagem mecanicista de Milkowski é similar. Ele se difere de Piccinini ao buscar uma explicação em termos de “processamento de informação”, de modo que mecanismos computacionais operem sobre estados que carreguem informações. Milkowski e Piccinini empregam suas respectivas teorias mecanicistas para defender o computacionalismo.

Computacionalistas mecanicistas tipicamente individualizam estados computacionais de maneira não semântica. Eles, portanto, encontram inquietações

sobre o papel explicativo do conteúdo representacional similares às encontradas pela CSF e pelo estruturalismo. Nesse sentido, Shagrir (2014) se queixa de que o computacionalismo mecanicista não acomoda as explicações das ciências cognitivas que são simultaneamente computacionais e representacionais. O reconhecimento da força dessa crítica irá depender da simpatia que se tenha pelo computacionalismo envolvido-de-conteúdo.

## 6.5 Pluralismo

Vimos várias concepções de computação contrastantes e que por vezes se sobrepõem: computação clássica, computação conexionista, computação neural, computação sintático-formal, computação envolvente-de-conteúdo, computação como processamento de informação, computação funcional, computação estruturalista e computação mecanicista. Cada concepção fornece uma forma diferente de computacionalismo. Cada concepção tem suas próprias forças e fraquezas. Pode-se adotar uma posição **pluralista** que reconheça concepções distintas como legítimas. Ao invés de elevar uma concepção acima das outras, pluralistas empregam alegremente a concepção que pareça útil em um dado contexto explicativo. Edelman (2008) segue uma linha pluralista, assim como o faz Chalmers (2012) em sua discussão mais recente.

A linha pluralista levanta algumas questões naturais. Podemos fornecer uma análise geral que englobe todos ou a maioria dos tipos de computação? Todas as computações compartilham umas com as outras certas características distintivas? Elas são, alternativamente, unidas por algo como uma semelhança de família? Um entendimento aprofundado da computação exige que enfrentemos essas questões.

## 7. Argumentos contra o computacionalismo

A TCM tem atraído muitas objeções. Em muitos casos, as objeções são aplicáveis apenas a versões específicas da TCM (tais como o computacionalismo clássico ou o computacionalismo conexionista). Aqui vão algumas objeções proeminentes. Veja também o verbete da SEP, **The Chinese room argument**, para uma discussão mais ampla acerca da objeção ao computacionalismo clássico desenvolvida por John Searle (1980).

## 7.1 Argumentos de trivialidade

Uma preocupação recorrente é a de que a TCM é **trivial**, pois podemos descrever quase qualquer sistema físico como executando computações. Searle (1990) alega que uma parede implementa **qualquer** programa de computador, desde que consigamos discernir algum padrão de movimentos de moléculas na parede que seja isomórfico à estrutura formal do programa. Putnam (1988, p. 121-125) defende uma tese acerca da trivialidade da TCM menos extrema, porém ainda muito forte, nessa mesma linha. Argumentos de trivialidade desempenham um papel significativo na literatura filosófica. Anticomputacionalistas empregam argumentos de trivialidade contra o computacionalismo, enquanto computacionalistas buscam evitar a trivialidade.

Computacionalistas normalmente rebatem os argumentos de trivialidade insistindo que os argumentos ignoram as restrições à implementação computacional, restrições que barram a trivialização das implementações. As restrições podem ser contrafactuais, causais, semânticas, ou de outro tipo, a depender da teoria da computação que se prefira. Por exemplo, David Chalmers (1995, 1996a) e B. Jack Copeland (1996) sustentam que o argumento da trivialidade de Putnam ignora condicionais contrafactuais que um sistema físico deve satisfazer para que implemente um modelo computacional. Outros filósofos dizem que um sistema físico deve ter propriedades representacionais para implementar um modelo computacional (FODOR, 1998, p. 11-12; LADYMAN, 2009; SPREVAK, 2010) ou pelo menos para implementar um modelo computacional envolvente-de-conteúdo (RESCORLA, 2013; 2014b). Os detalhes variam bastante e os computacionalistas debatem entre si sobre exatamente quais tipos de computação podem escapar de quais argumentos de trivialidade. Mas a maioria dos computacionalistas concordam que podemos escapar de quaisquer preocupações devastadoras envolvendo a trivialidade por meio de uma teoria suficientemente robusta acerca das relações de implementação entre os modelos computacionais e os sistemas físicos.

O *pancomputacionalismo* sustenta que cada sistema físico implementa um modelo computacional. Essa tese é plausível, uma vez que podemos argumentar que qualquer sistema físico implementa um modelo computacional suficientemente trivial (por exemplo, um autômato finito de um único estado). Como Chalmers (2011) notou, o pancomputacionalismo não parece preocupante para o computacionalismo.

O que seria preocupante seria uma tese de trivialidade muito mais forte a qual prescrevesse que quase todo o sistema físico implementa quase qualquer modelo computacional.

Para uma maior discussão sobre argumentos de trivialidade e implementação computacional, veja Sprevak (2019) e o verbete da SEP **computation in physical systems**.

## 7.2 Teorema da incompletude do Gödel

De acordo com alguns autores, os teoremas da incompletude de Gödel mostram que as capacidades matemáticas humanas superam as capacidades de qualquer máquina de Turing (NAGEL; NEWMAN, 1958). J. R. Lucas (1961) desenvolve essa posição em uma famosa crítica à TCCM. Roger Penrose desenvolve a crítica em *The Emperor's New Mind* (1989)<sup>27</sup> e escritos subsequentes. Vários filósofos e lógicos têm respondido à crítica, argumentando que as formulações existentes cometem falácias, petição de princípio e mesmo francos erros matemáticos (BOWIE, 1982; CHALMERS, 1996b; FEFERMAN, 1996; LEWIS, 1969; 1979; PUTNAM, 1975, p. 365-366; 1994; SHAPIRO, 2003). Há um amplo consenso de que esse criticismo à TCCM carece de qualquer força. Pode até acontecer de certas capacidades mentais humanas superarem a capacidade computacional de Turing, mas os teoremas da incompletude de Gödel não fornecem razões para antecipar esse resultado.

## 7.3 Limites da modelagem computacional

Um computador poderia compor a sinfonia *Eroica*? Ou descobrir a relatividade geral? Ou mesmo replicar a habilidade, que uma criança tem sem esforço algum, de perceber o ambiente, de amarrar seus cadarços e de discernir as emoções dos outros? Atividades humanas intuitivas, criativas ou habilidosas parecem resistir à formalização de um programa de computador (DREYFUS, 1972; 1992). De um modo mais geral, pode-se suspeitar que aspectos cruciais da cognição humana escapem à modelagem computacional, especialmente à modelagem computacional clássica.

---

<sup>27</sup> N.T.: *A Nova Mente do Imperador*, também traduzido para *A Mente Nova do Rei*.

Ironicamente, Fodor propaga uma versão contundente dessa crítica. Mesmo em suas formulações mais antigas da TCCM, Fodor (1975, p. 197-205) expressa um ceticismo considerável acerca da possibilidade de a TCCM lidar com todos os fenômenos cognitivos importantes. O pessimismo se torna mais notável em seus textos mais recentes (1983, 2000), os quais focam especialmente no *raciocínio abduutivo* como um fenômeno mental que potencialmente escapa à modelagem computacional. Seu argumento central pode ser resumido como:

- 1 A computação ao estilo Turing é sensível apenas a propriedades “locais” da representação mental, as quais são exauridas pela identidade e arranjo dos constituintes da representação.
- 2 Muitos processos mentais, paradigmaticamente a abdução, são sensíveis a propriedades “não locais” tais como relevância, simplicidade e conservadorismo.
- 3 Portanto, talvez precisemos abandonar a modelagem ao estilo Turing para os processos relevantes.
- 4 Infelizmente, atualmente não temos ideia de qual teoria alternativa possa servir como uma substituta adequada.

Alguns críticos negam (1), argumentando que computações ao estilo Turing adequadas podem ser sensíveis às propriedades “não locais” (SCHNEIDER, 2011; WILSON, 2005). Alguns desafiam (2), argumentando que inferências abdutivas típicas são sensíveis apenas às propriedades “locais” (CARRUTHERS, 2003; LUDWIG; SCHNEIDER, 2008; SPERBER, 2002). Alguns aceitam (3), mas disputam (4), insistindo que temos modelos promissores dos processos mentais relevantes que não são de estilo Turing (PINKER, 2005). Parcialmente estimulado por tais críticas, Fodor elaborou seu argumento em detalhes. Para defender (2), ele critica as teorias que modelam a abdução por meio do emprego de algoritmos heurísticos “locais” (2005, p. 41-46; 2008, p. 115-126) ou por meio da postulação de uma profusão de módulos cognitivos de domínios específicos (2005, p. 56-100). Para defender (4), ele critica várias teorias que lidam com a abdução por meio de modelos que não são ao estilo Turing (2000, p. 46-53; 2008), tais como as redes conexionistas.

O escopo e os limites da modelagem computacional permanecem como uma questão controversa. Nós podemos esperar que esse tópico continue sendo um foco ativo de investigação, examinado em conjunto com a IA.

## 7.4 Argumentos temporais

A atividade mental se desenvolve no tempo. Além disso, a mente realiza tarefas sofisticadas (por exemplo, estimativa perceptual) muito rapidamente. Muitos críticos suspeitam que o computacionalismo, especialmente o computacionalismo clássico, não acomoda adequadamente os aspectos temporais da cognição. Um modelo ao estilo Turing não faz menção explícita à escala de tempo na qual a computação ocorre. Pode-se implementar fisicamente a mesma máquina de Turing abstrata em um dispositivo à base de silício, ou em um dispositivo de tubos de vácuo mais lento, ou mesmo em um dispositivo de alavancas e roldanas ainda mais lento. Críticos recomendam que rejeitemos a TCCM em favor de algum enquadramento alternativo que incorpore mais diretamente as considerações temporais. Van Gelder e Port (1995) usam esse argumento para promover um **enquadramento** não computacional *de sistemas dinâmicos* para modelar a atividade mental. Eliasmith (2003; 2013, p. 12-13) o utiliza para sustentar sua Estruturação da Engenharia Neural.

Computacionalistas respondem que podemos **complementar** um modelo computacional abstrato com considerações temporais (PICCININI, 2010; WEISKOPF, 2004). Por exemplo, um modelo de máquina de Turing pressupõe “estágios de computação” discretos, sem descrever como os estágios se relacionam com o tempo físico. Mas podemos complementar nosso modelo descrevendo por quanto tempo cada estágio pode durar, desse modo, convertendo nosso modelo de máquina de Turing atemporal em uma teoria que provê predições temporais detalhadas. Muitos defensores da TCM empregam complementações nessa linha para estudar as propriedades temporais da cognição (NEWELL, 1990). Complementações similares figuram proeminentemente na ciência da computação, cujos praticantes estão muito preocupados em construir máquinas com as propriedades temporais apropriadas. Computacionalistas concluem que uma versão apropriadamente complementada da TCM pode capturar adequadamente como a cognição se desdobra no tempo.

Uma segunda objeção temporal enfatiza o contraste entre a evolução temporal *discreta* e *contínua* (VAN GELDER; PORT, 1995). A computação executada por uma máquina de Turing se desdobra em estágios discretos enquanto a atividade mental se desdobra em um tempo contínuo. Portanto, há uma profunda desconformidade entre as propriedades temporais da computação ao estilo Turing e aquelas da



atividade mental real. Precisamos de uma teoria psicológica que descreva a evolução temporal contínua.

Os computacionalistas respondem que essa objeção assume o que está tentando provar: que a atividade cognitiva não cai sob estágios discretos explicativamente significantes (WEISKOPF, 2004). Assumindo que o tempo físico é contínuo, segue-se que a atividade mental se desdobra em um tempo contínuo. **Não** se segue que os modelos cognitivos devam ter uma estrutura temporal contínua. Um computador pessoal opera em um tempo contínuo e seus estados físicos evoluem de forma contínua. Uma teoria física completa refletirá todas essas mudanças físicas. Mas nosso modelo **computacional** não reflete cada mudança física do computador. Nosso modelo computacional tem estrutura temporal discreta. Por que assumir que um bom modelo da mente ao nível cognitivo precisa refletir cada mudança física do cérebro? Mesmo que haja um *continuum* de estados **físicos** transcorrentes, por que assumir um *continuum* de estados **cognitivos** transcorrentes? O mero fato de a evolução temporal ser contínua não milita contra modelos computacionais com estruturas temporais discretas.

## 7.5 Cognição corporificada

A cognição corporificada é um programa de pesquisa que encontra inspiração no filósofo continental Maurice Merleau-Ponty, no psicólogo perceptual J. J. Gibson e em outras influências variadas. É um movimento bastante heterogêneo, mas a estratégia básica é enfatizar as ligações entre cognição, ação corporal e o ambiente circundante. Veja Varela, Thompson e Rosch (1991) para uma formulação inicial influente. Em muitos casos, os proponentes empregam ferramentas da teoria dos sistemas dinâmicos. Os proponentes normalmente apresentam sua abordagem como uma alternativa radical ao computacionalismo (CHEMERO, 2009; KELSO, 1995; THELEN; SMITH, 1994). A TCM, eles reclamam, trata a atividade mental como manipulação de símbolos estáticos afastada do ambiente circundante [*embedding environment*]. Ela negligencia uma miríade de maneiras complexas pelas quais o ambiente molda causal ou constitutivamente a atividade mental. Deveríamos substituir a TCM por um novo quadro que enfatize as ligações contínuas entre mente, corpo e ambiente. A dinâmica agente-ambiente, não a computação mental interna, é a chave para a compreensão da cognição. Frequentemente, uma atitude amplamente eliminativista em relação à intencionalidade impulsiona essa crítica.

Computacionalistas respondem que a TCM permite o devido reconhecimento da corporeidade da cognição. Modelos computacionais podem levar em consideração como mente, corpo e ambiente interagem continuamente. Afinal, modelos computacionais podem incorporar entradas sensoriais e saídas motoras. Não há razões óbvias para que uma ênfase na dinâmica agente-ambiente impeça uma ênfase também na computação mental interna (CLARK, 2014, p. 140-165; RUPERT, 2009). Computacionalistas defendem que a TCM pode incorporar quaisquer intuições legítimas oferecidas pelo movimento da cognição corporificada. Eles também insistem que a TCM continua sendo nosso melhor paradigma geral para explicar vários fenômenos psicológicos essenciais.

### Referência bibliográfica

- AITCHISON, L.; LENGYEL, M. The Hamiltonian Brain: Efficient Probabilistic Inference with Excitatory-Inhibitory Neural Circuit Dynamics. **PLoS Computational Biology**, v. 12, p. 1005186, 2016.
- ARJO, D. Sticking Up for Oedipus: Fodor on Intentional Generalizations and Broad Content. **Mind and Language**, v. 11, p. 231-245, 1996.
- AYDEDE, M. Fodor on Concepts and Frege Puzzles. **Pacific Philosophical Quarterly**, v. 79, p. 289-294, 1998.
- AYDEDE, M. Computationalism and Functionalism: Syntactic Theory of Mind Revisited. In: IRZIK, G.; GÜZELDERE, G. (Eds.). **Turkish Studies in the History and Philosophy of Science**. Dordrecht: Springer, 2005.
- AYDEDE, M.; ROBBINS, P. Are Frege Cases Exceptions to Intentional Generalizations?. **Canadian Journal of Philosophy**, v. 31, p. 1-22, 2001.
- BECHTEL, W.; ABRAHAMSEN, A. **Connectionism and the Mind**. Malden: Blackwell, 2002.
- BERMÚDEZ, J. **Philosophy of Psychology: A Contemporary Introduction**. New York: Routledge, 2005.
- BERMÚDEZ, J. **Cognitive Science: An Introduction to the Science of the Mind**. Cambridge: Cambridge University Press, 2010.
- BLOCK, N. Troubles With Functionalism. **Minnesota Studies in the Philosophy of Science**, v. 9, p. 261-325, 1978.
- BLOCK, N. Psychologism and Behaviorism. **Philosophical Review**, v. 90, p. 5-43, 1981.

- BLOCK, N. Mental Pictures and Cognitive Science. **Philosophical Review**, v. 92, p. 499-539, 1983.
- BLOCK, N. Advertisement for a Semantics for Psychology. **Midwest Studies in Philosophy**, v. 10, p. 615-678, 1986.
- BLOCK, N. Can the Mind Change the World?. In: BOOLOS, G. (Ed.). **Meaning and Method: Essays in Honor of Hilary Putnam**. Cambridge: Cambridge University Press, 1990.
- BLOCK, N. The Mind as the Software of the Brain. In: THINKING, E.; OSHERSON, B. (Eds.). **Invitation to Cognitive Science**. Cambridge, MA: MIT Press, 1995. v. 3.
- BLOCK, N.; FODOR, J. What Psychological States Are Not. **The Philosophical Review**, v. 81, p. 159-181, 1972.
- BODEN, M. Horses of a Different Color?. In: **RAMSEY et al.** (1991), p. 3-19.
- BONTLY, T. Individualism and the Nature of Syntactic States. **The British Journal for the Philosophy of Science**, v. 49, p. 557-574, 1998.
- BOWIE, G. Lucas's Number is Finally Up. **Journal of Philosophical Logic**, v. 11, p. 79-285, 1982.
- BROGAN, W. **Modern Control Theory**. Englewood Cliffs: Prentice Hall, 1990.
- BUCKNER, C. Deep Learning: A Philosophical Introduction. **Philosophy Compass**, v. 14, p. 12625, 2019.
- BUCKNER, C.; GARSON, J. **Connectionism and Post-Connectionist Models**. (Sprevak, Colombo, Eds.), 2019.
- BUESING, L. et al. Neural Dynamics of Sampling: A Model for Stochastic Computation in Recurring Networks of Spiking Neurons. **PLOS Computational Biology**, v. 7, p. 1002211, 2011.
- BURGE, T. Other Bodies. In: WOODFIELD, A. (Ed.). *Thought and Object*. Oxford: Oxford University Press, 1982. Reimpresso em BURGE (2007), p 82-89.
- BURGE, T. Individualism and Psychology. **The Philosophical Review**, v. 95, p. 3-45, 1986.
- BURGE, T. Individuation and Causation in Psychology. **Pacific Philosophical Quarterly**, v. 70, p. 303-322, 1989.
- BURGE, T. Intentional Properties and Causation. In: MACDONALD, C.; MACDONALD, G. (Eds.). **Philosophy of Psychology**. Oxford: Blackwell. Reimpresso em BURGE (2007), p. 334-343.
- BURGE, T. Disjunctivism and Perceptual Psychology. **Philosophical Topics**, v. 33, p. 1-78, 2005.
- BURGE, T. **Foundations of Mind**. Oxford: Oxford University Press, 2007.

- BURGE, T. **Origins of Objectivity**. Oxford: Oxford University Press, 2010a.
- BURGE, T. Origins of Perception. **Disputatio**, v. 4, p. 1-38, 2010b.
- BURGE, T. Steps Towards Origins of Propositional Thought. **Disputatio**, v. 4, p. 39-67, 2010c.
- BURGE, T. **Cognition through Understanding**. Oxford: Oxford University Press, 2013.
- CAMP, E. A Language of Baboon Thought?. In: LURZ, R. (Ed.). **The Philosophy of Animal Minds**. Cambridge: Cambridge University Press, 2009.
- CARRUTHERS, P. On Fodor's Problem. **Mind and Language**, v. 18, p. 508-523, 2003.
- CHALMERS, D. Syntactic Transformations on Distributed Representations. **Connection Science**, v. 2, p. 53-62, 1990.
- CHALMERS, D. Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation. **Philosophical Psychology**, v. 63, p. 305-319, 1993.
- CHALMERS, D. On Implementing a Computation. **Minds and Machines**, v. 4, p. 391-402, 1995.
- CHALMERS, D. Does a Rock Implement Every Finite State Automaton?. **Synthese**, v. 108, p. 309-333, 1996a.
- CHALMERS, D. Minds, Machines, and Mathematics. **Psyche**, v. 2, p. 11-20, 1996b.
- CHALMERS, D. The Components of Content. In: CHALMERS, D. (Ed.). **Philosophy of Mind: Classical and Contemporary Readings**. Oxford: Oxford University Press, 2002.
- CHALMERS, D. A Computational Foundation for the Study of Cognition. **The Journal of Cognitive Science**, v. 12, p. 323-357, 2011.
- CHALMERS, D. The Varieties of Computation: A Reply. **The Journal of Cognitive Science**, v. 13, p. 213-248, 2012.
- CHEMERO, A. **Radical Embodied Cognitive Science**. Cambridge, MA: MIT Press, 2009.
- CHENEY, D.; SEYFARTH, R. **Baboon Metaphysics: The Evolution of a Social Mind**. Chicago: University of Chicago Press, 2007.
- CHOMSKY, N. **Aspects of the Theory of Syntax**. Cambridge, MA: MIT Press, 1965.
- CHURCH, A. An Unsolvability Problem of Elementary Number Theory. **American Journal of Mathematics**, v. 58, p. 345-363, 1936.
- CHURCHLAND, P. Eliminative Materialism and the Propositional Attitudes. **Journal of Philosophy**, v. 78, p. 67-90, 1981.
- CHURCHLAND, P. **Neurophilosophy**. Cambridge, MA: MIT Press, 1986.
- CHURCHLAND, P. **A Neurocomputational Perspective: The Nature of Mind and the Structure of Science**. Cambridge, MA: MIT Press, 1989.

- CHURCHLAND, P. **The Engine of Reason, the Seat of the Soul**. Cambridge, MA: MIT Press, 1995.
- CHURCHLAND, P. **Neurophilosophy At Work**. Cambridge: Cambridge University Press, 2007.
- CHURCHLAND, P.; KOCH, C.; SEJNOWSKI, T. What Is Computational Neuroscience?. In: SCHWARTZ, E. (Ed.). **Computational Neuroscience**. Cambridge, MA: MIT Press, 1990.
- CHURCHLAND, P.; SEJNOWSKI, T. **The Computational Brain**. Cambridge, MA: MIT Press, 1992.
- CLARK, A. **Mindware: An Introduction to the Philosophy of Cognitive Science**. Oxford: Oxford University Press, 2014.
- CLAYTON, N.; EMERY, N.; DICKINSON, A. The Rationality of Animal Memory: Complex Caching Strategies of Western Scrub Jays. In: NUDDS, M.; HURLEY, S. (Eds.). **Rational Animals?** Oxford: Oxford University Press, 2006.
- COPELAND, J. What is Computation?. **Synthese**, v. 108, p. 335-359, 1996.
- COVER, T.; THOMAS, J. **Elements of Information Theory**. Hoboken: Wiley, 2006.
- CRANE, T. All the Difference in the World. **Philosophical Quarterly**, v. 41, p. 1-25, 1991.
- CRICK, F.; ASANUMA, C. Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex. In: MCCLELLAND et al (1987), p. 333-371.
- CUMMINS, R. **Meaning and Mental Representation**. Cambridge, MA: MIT Press, 1989.
- DAVIDSON, D. **Essays on Actions and Events**. Oxford: Clarendon Press, 1980.
- DAYAN, P. A Neurocomputational Jeremiad. **Nature Neuroscience**, v. 12, p. 1207, 2009.
- DENNETT, D. Intentional Systems. **Journal of Philosophy**, v. 68, p. 87-106, 1971.
- DENNETT, D. **The Intentional Stance**. Cambridge, MA: MIT Press, 1987.
- DENNETT, D. Mother Nature versus the Walking Encyclopedia. In: **RAMSEY et al** (1991), p. 21-30.
- DIETRICH, E. Semantics and the Computational Paradigm in Cognitive Psychology. **Synthese**, v. 79, p. 119-141, 1989.
- DONAHOE, J. Man as Machine: A Review of Memory and Computational Brain, by C.R. Gallistel and A.P. King, **Behavior and Philosophy**, v. 38, p. 83-101, 2010.
- DRETSKE, F. **Knowledge and the Flow of Information**. Oxford: Blackwell, 1981.
- DRETSKE, F. **Mental Events as Structuring Causes of Behavior**. Oxford: Clarendon Press, 1993.
- DREYFUS, H. **What Computers Can't Do**. Cambridge, MA: MIT Press, 1972.
- DREYFUS, H. **What Computers Still Can't Do**. Cambridge, MA: MIT Press, 1992.

- EDELMAN, S. **Computing the Mind**. Oxford: Oxford University Press, 2008.
- EDELMAN, S. How to Write a “How a Build a Brain” Book. **Trends in Cognitive Science**, v. 18, p. 118-119, 2014.
- EGAN, F. Must Psychology be Individualistic? **Philosophical Review**, v. 100, p. 179-203, 1991.
- EGAN, F. Individualism, Computation, and Perceptual Content. **Mind**, v. 101, p. 443-459, 1992.
- EGAN, F. In Defense of Narrow Mindedness. **Mind and Language**, v. 14, p. 177-194, 1999.
- EGAN, F. Naturalistic Inquiry: Where Does Mental Representation Fit In? In: ANTONY, L.; HOSTEIN, N. (Eds.). **Chomsky and His Critics**. Malden: Blackwell, 2003.
- EGAN, F. A Modest Role for Content. **Studies in History and Philosophy of Science**, v. 41, p. 253-259, 2010.
- EGAN, F. How to Think About Mental Content. **Philosophical Studies**, v. 170, p. 115-135, 2014.
- EGAN, F. The Nature and Function of Content in Computational Models. In: SPREVAK & COLOMBO (2019).
- ELIASMITH, C. Moving Beyond Metaphors: Understanding the Mind for What It Is. **Journal of Philosophy**, v. 100, p. 493-520, 2003.
- ELIASMITH, C. **How to Build a Brain**. Oxford: Oxford: University Press, 2013.
- ELIASMITH, C.; ANDERSON, C. **Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems**. Cambridge, MA: MIT Press, 2003.
- ELMAN, J. Finding Structure in Time. **Cognitive Science**, v. 14, p. 179-211, 1990.
- FEFERMAN, S. Penrose’s Gödelian Argument. **Psyche**, v. 2, p. 21-32, 1996.
- FELDMAN, J.; BALLARD, D. Connectionist Models and their Properties. **Cognitive Science**, v. 6, p. 205-254, 1982.
- FIELD, H. **Truth and the Absence of Fact**. Oxford: Clarendon Press, 2001.
- FIGDOR, C. Semantic Externalism and the Mechanics of Thought. **Minds and Machines**, v. 19, p. 1-24, 2009.
- FODOR, J. **The Language of Thought**. New York: Thomas Y. Crowell, 1975.
- FODOR, J. Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. **Behavioral and Brain Science**, v. 3, p. 63-73, 1980.
- FODOR, J. **Representations**. Cambridge: MIT Press, 1981.
- FODOR, J. **The Modularity of Mind**. Cambridge, MA: MIT Press, 1983.
- FODOR, J. **Psychosemantics**. Cambridge: MIT Press, 1987.
- FODOR, J. **A Theory of Content and Other Essays**. Cambridge, MA: MIT Press, 1990.

- FODOR, J. A Modal Argument for Narrow Content. **Journal of Philosophy**, v. 88, p. 5-26, 1991.
- FODOR, J. **The Elm and the Expert**. Cambridge, MA: MIT Press, 1994.
- FODOR, J. **Concepts**. Oxford: Clarendon Press, 1998.
- FODOR, J. **The Mind Doesn't Work That Way**. Cambridge, MA: MIT Press, 2000.
- FODOR, J. Reply to Steven Pinker "So How Does the Mind Work?". **Mind and Language**, v. 20, p. 25-32, 2005.
- FODOR, J. **LOT2**. Oxford: Clarendon Press, 2008.
- FODOR, J.; PYLYSHYN, Z. Connectionism and Cognitive Architecture: A Critical Analysis. **Cognition**, v. 28, p. 3-71, 1988.
- FREGE, G. Begriffsschrift, eine der Arithmetischen Nachgebildete Formelsprache des Reinen Denkens. Reimpresso como: Concept Script, a Formal Language of Pure Thought Modeled upon that of Arithmetic. In: VAN HEIJENOORT (Ed.). **From Frege to Gödel: A Source Book in Mathematical Logic**. Tradução: S. Bauer-Mengelberg. Cambridge: Harvard University Press, 1978/1967.
- GALLISTEL, C. **The Organization of Learning**. Cambridge, MA: MIT Press, 1990.
- GALLISTEL, C.; KING, A. **Memory and the Computational Brain**. Malden: Wiley-Blackwell, 2009.
- GANDY, R. Church's Thesis and Principles for Mechanism. In: BARWISE, J.; KEISLER, H.; KUNEN, K. (Eds.). *The Kleene Symposium*. Amsterdam: North Holland, 1980.
- GÖDEL, K. On Formally Undecidable Propositions of Principia Mathematica and Related Systems. Reimpresso com um Postscript em DAVIS, M. (Ed.). **The Undecidable**. New York: Raven Press Books, 1936/65.
- GRICE, P. **Studies in the Ways of Words**. Cambridge: Harvard University Press, 1989.
- HADLEY, R. Cognition and the Computational Power of Connectionist Networks. **Connection Science**, v. 12, p. 95-110, 2000.
- HARNISH, R. **Minds, Brains, Computers**. Malden: Blackwell, 2002.
- HAUGELAND, J. **Artificial Intelligence: The Very Idea**. Cambridge, MA: MIT Press, 1985.
- HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. New York: Prentice Hall, 2008.
- HORGAN, T.; TIENSON, J. **Connectionism and the Philosophy of Psychology**. Cambridge, MA: MIT Press, 1996.
- HOROWITZ, A. Computation, External Factors, and Cognitive Explanations. **Philosophical Psychology**, v. 20, p. 65-80, 2007.

- JOHNSON, K. On the Systematicity of Language and Thought. **Journal of Philosophy**, v. 101, p. 111-139, 2004.
- JOHNSON-LAIRD, P. **The Computer and the Mind**. Cambridge: Harvard University Press, 1988.
- JOHNSON-LAIRD, P. The History of Mental Models. In: MANKTELOW, K.; CHUNG, M. (Eds.). **Psychology of Reasoning: Theoretical and Historical Perspectives**. New York: Psychology Press, 2004.
- KAZEZ, J. Computationalism and the Causal Role of Content. **Philosophical Studies**, v. 75, p. 231-260, 1995.
- KELSO, J. **Dynamic Patterns**. Cambridge, MA: MIT Press, 1995.
- KLEIN, C. Two Paradigms for Individuating Implementations. **Journal of Cognitive Science**, v. 13, p. 167-179, 2012.
- KRIEGESGORTE, K. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. **Annual Review of Vision Science**, v. 1, p. 417-446, 2015.
- KRIEGESGORTE, K.; DOUGLAS, P. Cognitive Computational Neuroscience. **Nature Neuroscience**, v. 21, p. 1148-1160, 2018.
- KRISHEVSKY, A.; SUTSKEVER, I.; HINTON, G. ImageNet Classification with Deep Convolutional Neural Networks. **Advances in Neural Information Processing Systems**, v. 25, p. 1097-1105, 2012.
- KROTOV, D.; HOPFIELD, J. Unsupervised Learning by Competing Hidden Units. **Proceedings of the National Academy of Sciences**, v. 116, p. 7723-7731, 2019.
- LADYMAN, J. What Does it Mean to Say that a Physical System Implements a Computation? **Theoretical Computer Science**, v. 410, p. 376-383, 2009.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. **Nature**, v. 521, p. 436-444, 2015.
- LEWIS, D. Lucas against Mechanism. **Philosophy**, v. 44, p. 231-3, 1969.
- LEWIS, D. Psychophysical and Theoretical Identifications. **Australasian Journal of Philosophy**, v. 50, p. 249-58, 1972.
- LEWIS, D. Lucas Against Mechanism II. **Canadian Journal of Philosophy**, v. 9, p. 373-376, 1979.
- LEWIS, D. Reduction of Mind. In: GUTTENPLAN, S. (Ed.). **A Companion to the Philosophy of Mind**. Oxford: Blackwell, 1994.
- LIZIER, J.; FLECKER, B.; WILLIAMS, P. Towards a Synergy-based Account of Measuring Information Modification. In: **Proceedings of the 2013 IEEE Symposium on Artificial Life**. Singapore, 2013, p. 43-51.



- LUCAS, J. Minds, Machines, and Gödel. **Philosophy**, v. 36, p. 112-137, 1961.
- LUDWIG, K.; SCHNEIDER, S. Fodor's Critique of the Classical Computational Theory of Mind. **Mind and Language**, v. 23, p. 123-143, 2008.
- MA, W. Bayesian Decision Models: A Primer. **Neuron**, v. 104, p. 164-175, 2019.
- MAASS, W. Networks of Spiking Neurons: The Next Generation of Neural Network Models. **Neural Networks**, v. 10, p. 1659-1671, 1997.
- MACLENNAN, B. Analog Computation. In: MEYERS, R. (Ed.). **Computational Complexity**. New York: Springer, 2012.
- MARBLESTONE, A.; WAYNE, G.; KORDING, K. Toward an Integration of Deep Learning and Neuroscience. **Frontiers in Computational Neuroscience**, v. 10, p. 1-41, 2016.
- MARCUS, G. **The Algebraic Mind**. Cambridge, MA: MIT Press, 2001.
- MARR, D. **Vision**. San Francisco: W.H. Freeman, 1982.
- MCCLELLAND, J.; RUMELHART, D.; HINTON, G. The Appeal of Parallel Distributed Processing. In: RUMELHART et al (1986), p. 3-44.
- MCCLELLAND, J.; RUMELHART, D.; PDP Research GROUP. **Parallel Distributed Processing**. Cambridge, MA: MIT Press, 1987.
- MCCULLOCH, W.; PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. **Bulletin of Mathematical Biophysics**, v. 7, p. 115-133, 1943.
- MCDERMOTT, D. **Mind and Mechanism**. Cambridge, MA: MIT Press, 2001.
- MENDOLA, J. **Anti-Externalism**. Oxford: Oxford University Press, 2008.
- MILKOWSKI, M. **Explaining the Computational Mind**. Cambridge, MA: MIT Press, 2013.
- MILLER, P. **An Introductory Course in Computational Neuroscience**. Cambridge, MA: MIT Press, 2018.
- MOLE, C. Dead Reckoning in the Desert Ant: A Defense of Connectionist Models. **Review of Philosophy and Psychology**, v. 5, p. 277-290, 2014.
- MURPHY, K. **Machine Learning: A Probabilistic Perspective**. Cambridge, MA: MIT Press, 2012.
- NAGEL, E.; NEWMAN, J. **Gödel's Proof**. New York: New York University Press, 1958.
- NASELARIS, T. et al. Cognitive Computational Neuroscience: A New Conference for an Emerging Discipline. **Trends in Cognitive Science**, v. 22, p. 365-367, 2018.
- NEWELL, A. Computer Science as Empirical Inquiry: Symbols and Search. **Communications of the ACM**, v. 19, p. 113-126, 1976.
- NEWELL, A. **Unified Theories of Cognition**. Cambridge: Harvard University Press, 1990.

- NEWELL, A.; SIMON, H. The Logic Theory Machine: A Complex Information Processing System. **IRE Transactions on Information Theory**, v. IT-2, 3, p. 61-79, 1956.
- OCKHAM, W. **Summa Logicae**, in his **Philosophical Writings, A Selection**. London: Nelson, 1957.
- O'KEEFE, J.; NADEL, L. **The Hippocampus as a Cognitive Map**. Oxford: Clarendon University Press, 1978.
- ORHAN, A.; MA, W. Efficient Probabilistic Inference in Generic Neural Networks Trained with Non-probabilistic Feedback. **Nature Communications**, v. 8, p. 1-14, 2017.
- PEACOCKE, C. **A Study of Concepts**. Cambridge, MA: MIT Press, 1992.
- PEACOCKE, C. Externalist Explanation. **Proceedings of the Aristotelian Society**, v. 67, p. 203-230, 1993.
- PEACOCKE, C. Content, Computation, and Externalism. **Mind and Language**, v. 9, p. 303-335, 1994.
- PEACOCKE, C. Computation as Involving Content: A Response to Egan. **Mind and Language**, v. 14, p. 195-202, 1999.
- PENROSE, R. **The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics**. Oxford: Oxford University Press, 1989.
- PERRY, J. Broadening the Mind. **Philosophy and Phenomenological Research**, v. 58, p. 223-231, 1998.
- PIANTADOSI, S.; TENENBAUM, J.; GOODMAN, N. Bootstrapping in a Language of Thought. **Cognition**, v. 123, p. 199-217, 2012.
- PICCININI, G. Functionalism, Computationalism, and Mental States. **Studies in History and Philosophy of Science**, v. 35, p. 811-833, 2004.
- PICCININI, G. Computing Mechanisms. **Philosophy of Science**, v. 74, p. 501-526, 2007.
- PICCININI, G. Computation Without Representation. **Philosophical Studies**, v. 137, p. 205-241, 2008a.
- PICCININI, G. Some Neural Networks Compute, Others Don't. **Neural Networks**, v. 21, p. 311-321, 2008b.
- PICCININI, G. The Resilience of Computationalism. **Philosophy of Science**, v. 77, p. 852-861, 2010.
- PICCININI, G. Computationalism. In: MARGOLIS, E.; SAMUELS, R.; STICH, S. (Eds.). **The Oxford Handbook of Philosophy and Cognitive Science**. Oxford: Oxford University Press, 2012.

- PICCININI, G. **Physical Computation: A Mechanistic Account**. Oxford: Oxford University Press, 2015.
- PICCININI, G.; BAHAR, S. Neural Computation and the Computational Theory of Cognition. **Cognitive Science**, v. 37, p. 453-488, 2013.
- PICCININI, G.; SCARANTINO, A. Computation vs. Information processing: Why their Difference Matters to Cognitive Science. **Studies in History and Philosophy of Science**, v. 41, p. 237-246, 2010.
- PICCININI, G.; SHAGRIR, O. Foundations of Computational Neuroscience. **Current Opinion in Neurobiology**, v. 25, p. 25-30, 2014.
- PINKER, S. So How Does the Mind Work?. **Mind and Language**, v. 20, p. 1-24, 2005.
- PINKER, S.; PRINCE, A. On Language and Connectionism. **Cognition**, v. 28, p. 73-193, 1988.
- POUGET, A. et al. Probabilistic Brains: Knowns and Unknowns. **Nature Neuroscience**, v. 16, p. 1170-1178, 2013.
- PUTNAM, H. Psychophysical Predicates. In: CAPITAN, W; MERRILL, D. (Eds.). **Art, Mind, and Religion**. Pittsburgh: University of Pittsburgh Press, 1967. Reimpresso em PUTNAM (1975) como "The Nature of Mental States", p. 429-440.
- PUTNAM, H. **Mind, Language, and Reality: Philosophical Papers**. Cambridge: Cambridge University Press, 1975. v. 2
- PUTNAM, H. **Realism and Reason: Philosophical Papers**. Cambridge: Cambridge University Press, 1983. v. 3
- PUTNAM, H. **Representation and Reality**. Cambridge, MA: MIT Press, 1988.
- PUTNAM, H. The Best of All Possible Brains? **The New York Times**, p. 7, 1994.
- PYLYSHYN, Z. **Computation and Cognition**. Cambridge, MA: MIT Press, 1984.
- QUINE, W. **Word and Object**. Cambridge, MA: MIT Press, 1960.
- RAMSEY, W.; STICH, S.; RUMELHART. **Philosophy and Connectionist Theory**. Hillsdale: Lawrence Erlbaum Associates, 1991.
- RESCORLA, M. Chrysippus's Dog as a Case Study in Non-Linguistic Cognition. In: LURZ, R. (Ed.). **The Philosophy of Animal Minds**. Cambridge: Cambridge University Press, 2009a.
- RESCORLA, M. Cognitive Maps and the Language of Thought. **The British Journal for the Philosophy of Science**, v. 60, p. 377-407, 2009b.
- RESCORLA, M. How to Integrate Representation into Computational Modeling, and Why We Should. **Journal of Cognitive Science**, v. 13, p. 1-38, 2012.

- RESCORLA, M. Against Structuralist Theories of Computational Implementation. **British Journal for the Philosophy of Science**, v. 64, p. 681-707, 2013.
- RESCORLA, M. The Causal Relevance of Content to Computation. **Philosophy and Phenomenological Research**, v. 88, p. 173-208, 2014a.
- RESCORLA, M. A Theory of Computational Implementation. **Synthese**, v. 191, p. 1277-1307, 2014b.
- RESCORLA, M. Bayesian Perceptual Psychology. In: MATTHEN, M. (Ed.). **The Oxford Handbook of the Philosophy of Perception**. Oxford: Oxford University Press, 2015.
- RESCORLA, M. From Ockham to Turing—and Back Again. In BOKULICH, A.; FLOYD, J. (Eds.). **Turing 100: Philosophical Explorations of the Legacy of Alan Turing, (Boston Studies in the Philosophy and History)**. Springer, 2017a.
- RESCORLA, M. Levels of Computational Explanation”. In: POWERS, T. (Ed.). **Philosophy and Computing: Essays in Epistemology, Philosophy of Mind, Logic, and Ethics**. Cham: Springer, 2017b.
- RESCORLA, M. A Realist Perspective on Bayesian Cognitive Science. In: NES, A.; CHAN, T. (Eds.). **Inference and Consciousness**. New York: Routledge, 2020.
- ROGERS, T.; MCCLELLAND, J. Parallel Distributed Processing at 25: Further Explorations of the Microstructure of Cognition. **Cognitive Science**, v. 38, p. 1024-1077, 2014.
- RUMELHART, D. The Architecture of Mind: A Connectionist Approach. In: POSNER, M. (Ed.). **Foundations of Cognitive Science**. Cambridge, MA: MIT Press, 1989.
- RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning Representations by Back-propagating Errors. **Nature**, v. 323, p. 533-536, 1986.
- RUMELHART, D.; MCCLELLAND, J. PDP Models and General Issues in Cognitive Science. In: RUMELHART et al (1986), p. 110-146.
- RUMELHART, D.; MCCLELLAND, J.; PDP RESEARCH GROUP. **Parallel Distributed Processing**. Cambridge: MIT PRESS, 1986.
- RUPERT, R. Frege's Puzzle and Frege Cases: Defending a Quasi-Syntactic Solution. **Cognitive Systems Research**, v. 9, p. 76-91, 2008.
- RUPERT, R. **Cognitive Systems and the Extended Mind**. Oxford: Oxford University Press, 2009.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. New York: Prentice Hall, 2010.

- SAWYER, S. There Is No Viable Notion of Narrow Content. In: MCLAUGHLIN, B.; COHEN, J. (Eds.). **Contemporary Debates in Philosophy of Mind**. Malden: Blackwell, 2000.
- SCHNEIDER, S. Direct Reference, Psychological Explanation, and Frege Cases. **Mind and Language**, v. 20, p. 423-447, 2005.
- SCHNEIDER, S. **The Language of Thought: A New Philosophical Direction**. Cambridge, MA: MIT Press, 2011.
- SEARLE, J. Minds, Brains, and Programs. **Behavioral and Brain Sciences**, v. 3, p. 417-457, 1980.
- SEARLE, J. Is the Brain a Digital Computer? **Proceedings and Addresses of the American Philosophical Association**, v. 64, p. 21-37, 1990.
- SEGAL, G. **A Slim Book About Narrow Content**. Cambridge, MA: MIT Press, 2000.
- SHAGRIR, O. Content, Computation, and Externalism. **Mind**, v. 110, p. 369-400, 2001.
- SHAGRIR, O. Why We View the Brain as a Computer. **Synthese**, v. 153, p. 393-416, 2006.
- SHAGRIR, O. Review of Explaining the Computational Theory of Mind, by Marcin Milkowski. **Notre Dame Review of Philosophy**, 2014.
- SHAGRIR, O. In Defense of the Semantic View of Computation. **Synthese**, v. 197, n. 9, p. 4083-4108, 2020.
- SHANNON, C. A Mathematical Theory of Communication. **Bell System Technical Journal**, v. 27, p. 379-423, 623-656, 1948.
- SHAPIRO, S. Truth, Mechanism, and Penrose's New Argument. **Journal of Philosophical Logic**, v. 32, p. 19-42, 2003.
- SHEA, N. Naturalizing Representational Content. **Philosophy Compass**, v. 8, p. 496-509, 2013.
- SHEA, N. **Representation in Cognitive Science**. Oxford: Oxford University Press, 2018.
- SIEG, W. On Computability. In: IRVINE, A. (Ed.). **Philosophy of Mathematics**. Burlington: Elsevier, 2009.
- SIEGELMANN, H.; SONTAG, E. Turing Computability with Neural Nets. **Applied Mathematics Letters**, v. 4, p. 77-80, 1991.
- SIEGELMANN, H.; SONTAG, E. On the Computational Power of Neural Nets. **Journal of Computer and Science Systems**, v. 50, p. 132-150, 1995.
- SILVERBERG, A. Chomsky and Egan on Computational Theories of Vision. **Minds and Machines**, v. 16, p. 495-524, 2006.
- SLOMAN, A. **The Computer Revolution in Philosophy**. Hassocks: The Harvester Press, 1978.

- SMOLENSKY, P. On the Proper Treatment of Connectionism. **Behavioral and Brain Sciences**, v. 11, p. 1-74, 1988.
- SMOLENSKY, P. Connectionism, Constituency, and the Language of Thought. In: FODOR; HIS CRITICS, B.; REY, G. (Eds.). **Meaning in Mind**. Cambridge: Blackwell, 1991.
- SPERBER, D. In Defense of Massive Modularity. In: DUPOUX, E. (Ed.). **Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler**. Cambridge, MA: MIT Press, 2002.
- SPREVAK, M. Computation, Individuation, and the Received View on Representation. **Studies in History and Philosophy of Science**, v. 41, p. 260-270, 2010.
- SPREVAK, M. Triviality Arguments About Computational Implementation. In: SPREVAK & COLOMBO (2019).
- SPREVAK, M. Two Kinds of Information Processing in Cognition. **Review of Philosophy and Psychology**, v. 11, p. 591-611, 2020.
- SPREVAK, M.; COLOMBO, M. (Eds.). **The Routledge Handbook of the Computational Mind**. New York: Routledge, 2019.
- STALNAKER, R. **Context and Content**. Oxford: Oxford University Press, 1999.
- STICH, S. **From Folk Psychology to Cognitive Science**. Cambridge, MA: MIT Press, 1983.
- THELEN, E.; SMITH, L. **A Dynamical Systems Approach to the Development of Cognition and Action**. Cambridge, MA: MIT Press, 1994.
- THRUN, S.; BURGARD, W.; FOX, D. **Probabilistic Robotics**. Cambridge, MA: MIT Press, 2006.
- THRUN, S.; MONTEMERLO, M.; DAHLKAMP, H. Stanley: The Robot That Won the DARPA Grand Challenge. **Journal of Field Robotics**, v. 23, p. 661-692, 2006.
- TOLMAN, E. Cognitive Maps in Rats and Men. **Psychological Review**, v. 55, p. 189-208, 1948.
- TRAPPENBERG, T. **Fundamentals of Computational Neuroscience**. Oxford: Oxford University Press, 2010.
- TURING, A. On Computable Numbers, with an Application to the Entscheidungsproblem. **Proceedings of the London Mathematical Society**, v. 42, p. 230-265, 1936.
- TURING, A. Computing Machinery and Intelligence. **Mind**, v. 49, p. 433-460, 1950.
- VAN GELDER, T. Compositionality: A Connectionist Variation on a Classical Theme. **Cognitive Science**, v. 14, p. 355-384, 1990.
- VAN GELDER, T.; PORT, R. It's About Time: An Overview of the Dynamical Approach to Cognition. In: PORT, R.; VAN GELDER, T. (Eds.). **Mind as Motion:**

- Explorations in the Dynamics of Cognition.** Cambridge, MA: MIT Press, 1995.
- VARELA, F.; THOMPSON, E.; ROSCH, E. **The Embodied Mind: Cognitive Science and Human Experience.** Cambridge, MA: MIT Press, 1991.
- VON NEUMANN, J. First Draft of a Report on the EDVAC. Philadelphia, PA: Moore School of Electrical Engineering, University of Pennsylvania, 1945.
- WAKEFIELD, J. Broad versus Narrow Content in the Explanation of Action: Fodor on Frege Cases. **Philosophical Psychology**, v. 15, p. 119-133, 2002.
- WEISKOPF, D. The Place of Time in Cognition. **British Journal for the Philosophy of Science**, v. 55, p. 87-105, 2004.
- WHITEHEAD, A.; RUSSELL, B. **Principia Mathematica**, v. 1, 1925.
- WILSON, R. What Computers (Still, Still) Can't Do. **New Essays in Philosophy of Language and Mind**, v. 30, p. 407-425, 2005.
- YABLO, S. Wide Causation. **Philosophical Perspectives**, v. 11, p. 251-281, 1997.
- YABLO, S. Causal Relevance. **Philosophical Issues**, v. 13, p. 316-327, 2003.
- ZEDNIK, C. **Computational Cognitive Neuroscience** (Sprevak, Colombo, Eds.), 2019.
- ZYLBERBERG, A. et al. The Human Turing Machine. **Trends in Cognitive Science**, v. 15, p. 293-300, 2011.

## (III) Teorias Teleológicas do Conteúdo Mental\*

Autora: Karen Neander

Tradução: Jeferson Huffermann (PPGFIL-UFRGS)

Revisão: Nara Figueiredo (UNICAMP)

César Fernando Meurer (UFABC)

Teorias teleológicas do conteúdo mental tentam explicar os conteúdos de representações mentais fazendo apelo à noção teleológica de função. Considere, por exemplo, o pensamento que arbustos estão florescendo. Numa teoria representacional do pensamento, ele envolve uma representação dos arbustos florescendo. Uma teoria do conteúdo almeja, entre outras coisas, nos dizer por que esta representação tem esse conteúdo; almeja dizer por que é um pensamento sobre o florescimento dos arbustos em vez de ser sobre o sol a brilhar ou sobre porcos voando ou sobre absolutamente nada. Em geral, uma teoria do conteúdo tenta dizer por que uma representação representa aquilo que representa.

---

\* NEANDER, K. Teleological Theories of Mental Content. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2018. Disponível em: <https://plato.stanford.edu/archives/spr2018/entries/content-teleological/>. Acesso em: 10 out. 2021.

The following is the translation of the entry on Teleological Theories of Mental Content by Karen Neander, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/spr2018/entries/content-teleological/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/content-teleological/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.



De acordo com teorias teleológicas do conteúdo, o que uma representação representa depende das funções dos sistemas que produzem ou usam a representação. É dito que a noção relevante de representação é aquela utilizada em biologia e neurobiologia ao atribuir funções a componentes de organismos (como em “a função da glândula pineal é secretar melatonina” e “a função da área cerebral MT é processar informação sobre movimento”). Proponentes de teorias teleológicas do conteúdo geralmente entendem tais funções como sendo o que a coisa com aquela função foi selecionada para, seja por seleção natural (como comumente entendida), ou por algum outro processo natural de seleção.

## 1. Objetivos gerais

Muitos estados mentais (talvez todos) são sobre coisas ou são direcionados às coisas, assim como a crença que a primavera se aproxima é sobre a primavera se aproximar ou como o desejo por chocolate é direcionado ao chocolate. O filósofo Franz Brentano (1838-1917) falou de tais estados mentais como envolvendo apresentações dos objetos dos nossos pensamentos. A ideia era que não poderíamos desejar chocolate a não ser que chocolate estivesse de algum modo presente em nossas mentes. Hoje em dia, diríamos que chocolate deve ser representado em nossas mentes se é chocolate o que desejamos. Teorias teleológicas do conteúdo, como outras teorias do conteúdo mental, tentam resolver o que se refere comumente como problema de Brentano: o problema em explicar intencionalidade, explicar como estados mentais podem ser sobre coisas ou direcionados a coisas desse modo.

Uma versão do problema muitas vezes atribuída a Brentano, mas possivelmente mais corretamente atribuível a Roderick Chisholm (1957), concerne pensamentos sobre objetos não existentes. Chisholm argumentou que a *acerca-de*<sup>28</sup> (ou intencionalidade) de estados mentais não pode ser uma relação física entre o estado mental e aquilo sobre o que ele é (seu objeto) porque em uma relação física cada um dos *relata* deve existir, mas os objetos de estados mentais podem não existir. Se André beija Cátia, ambos André e Cátia devem existir, e se o sol brilha sobre o jardim, ambos o jardim e o sol também devem existir. De modo contrastante, Guilherme pode amar o Papai Noel e procurar unicórnios, mesmo que ambos não existam.

---

<sup>28</sup> N.T.: *Aboutness*

Chisholm concluiu que é difícil ver como intencionalidade pode ser um fenômeno físico, mas aqueles que nos oferecem teorias teleológicas (quase sempre) adotam um enquadramento fisicista enquanto tentam explicar como a intencionalidade é possível. Almejam o que seguidamente chama-se “teoria naturalista”: “naturalista” porque almeja prover uma teoria que é consistente com a tese de que a mobília fundamental do universo em nada difere do que é descrito pelas ciências naturais. Dentro deste enquadramento, uma hipótese de trabalho é que intencionalidade não é ontologicamente fundamental, então a maioria das teorias teleológicas tentam mostrar que a intencionalidade é parte do mundo natural ao mostrar como pode ser entendida em termos de outros entes naturais. Consequentemente, proponentes de teorias teleológicas do conteúdo mental tentam dizer qual o motivo de uma representação mental, *R*, representa o que representa, *C*, preenchendo o espaço vazio em “*R* tem o conteúdo *C* porque (em virtude de) \_\_\_\_\_” sem fazer uso de termos intencionais inelimináveis.

Essas teorias também precisam dar conta da natureza normativa de representações mentais. O conteúdo é dito normativo porque legitima certas avaliações. Avaliamos crenças como verdadeiras ou falsas, memórias como acuradas ou inacuradas, percepções como verídicas ou ilusórias, e assim por diante. Também avaliamos desejos como satisfeitos ou não, e instruções motoras como corretamente ou incorretamente executadas. O conteúdo que é normativo ocasionalmente é descrito como *vero-valorável*<sup>29</sup>. Estados representacionais são considerados verdadeiros ou falsos tanto em virtude dos estados do mundo, quanto por seu conteúdo. Por exemplo, a verdade da minha crença que hoje está ensolarado depende de o dia estar ensolarado, mas também é dependente de ser uma crença de que hoje está ensolarado. Se o conteúdo da minha crença fosse diferente (por exemplo, se eu acreditasse que hoje está quente), o valor de verdade poderia ser diferente. A natureza normativa do conteúdo impõe um problema para teorias naturalistas, mas proponentes de teorias teleológicas do conteúdo mental pensam que esse problema é tratável.

---

<sup>29</sup> N.T.: *Truth-evaluable*

Muita atenção é dada à possibilidade de deturpação<sup>30</sup> representacional. Isso porque a distinção entre representação correta e incorreta frequentemente é considerada como uma distinção normativa central, além disso, a capacidade de deturpar é pensada como essencial ao representar: se não há a possibilidade de deturpar, não há representação. Considere a representação mental de um gato. Se deve ter o conteúdo **gato**, de modo que todos os gatos (e apenas eles) incluam-se em sua extensão, deve ser o caso de que, se fosse usada para classificar um não-gato (por exemplo, um cão) isso seria deturpação. Entretanto, pode haver exceções à regra de que todas as representações podem deturpar (por exemplo, a representação que tem como seu conteúdo **algo ou nada**). Deturpações também não são possíveis em todo tipo de contexto mental (por exemplo, em sonhos e, talvez, desejos).

A possibilidade de deturpação também se conecta com a preocupação de Chisholm com objetos inexistentes, porque a capacidade de deturpar equivale à capacidade básica de representar objetos não existentes. Imagine um dispositivo simples de detecção que normalmente entra em um estado VERMELHO em resposta ao vermelho. Se o estado VERMELHO tem o conteúdo “ali é vermelho”, e se VERMELHO pudesse ser às vezes instanciado quando nada vermelho está presente, uma instância de VERMELHO poderia representar uma instância não existente de vermelho. Há mais para explicar em nossa capacidade de representar objetos inexistentes do que explicar como deturpações são possíveis, mas explicar como são possíveis é um ponto de partida.

Deturpações tornam claro que representar é muitas vezes uma tríade. Suponha que, por exemplo, numa noite escura vejo alguns jornais amassados soprados pelo vento como um gato esgueirando-se pela rua. Há ao menos três coisas envolvidas. Primeiro, há a representação (ou veículo da representação) que tem um conteúdo. Em nós, presumivelmente algum tipo de estado ou evento neuronal. Aqui, tais representações mentais são denotadas pela expressão em português capitalizada (por exemplo, GATO). Segundo, há aquilo que a representação

---

<sup>30</sup> N.T.: o termo original *misrepresentation* será traduzido como “deturpação”, mas também poderia ser traduzido como “representação errônea” ou “representação falsa”. É o oposto de uma representação que representa adequadamente seu alvo. O termo “deturpação” em muitos contextos de uso tem como conotação uma intenção enganadora ou maliciosa. Nosso uso não tem conotações morais desse tipo.

visa representar, nesse caso o jornal. Cummins (1996) chama isso de *alvo* da representação. E terceiro, há o conteúdo da representação. Visto que represento o jornal como um gato, o conteúdo da representação nesse caso é *gato*. Nesse caso, ocorreu uma deturpação porque o alvo da representação não é a extensão da representação; o jornal não é um gato.

Podemos perguntar sobre cada uma dessas três posições na relação representacional. Primeiro, a questão do *status* representacional: Por que GATO conta como um caso de representação? Ou, de modo mais geral, qual é a diferença entre estados naturais que são representacionais e estados que não são? Segundo, a questão de determinar o alvo: o que faz com que uma instância de GATO tenha o jornal como seu alvo? Terceiro, a questão da determinação do conteúdo: o que faz ser o caso que GATO tenha o conteúdo **gato**? Ou, de modo mais geral, em virtude do que qualquer representação tem o conteúdo que tem? Teorias teleológicas do conteúdo mental estão primariamente preocupadas com a determinação do conteúdo, mas uma solução completa do problema de Brentano dará respostas para todos os três questionamentos.

Uma distinção às vezes feita é entre representação **de** e representação **como**. Se teorias teleológicas estão preocupadas ou não com representações **de** ou representações **como** depende de como essas locuções são usadas. Num certo sentido, referindo-se ao exemplo prévio, minha representação GATO representa o jornal **como** um gato, mesmo sendo uma representação *de* um jornal. Nessa maneira de falar, teorias teleológicas do conteúdo são teorias de representações **como**. Entretanto, as palavras “como” e “de” não são sempre diagnósticas da distinção conteúdo/alvo. Por exemplo, podemos também dizer que usamos a representação **de** um gato para representar o jornal.

As teorias teleológicas que estão correntemente no mercado são geralmente teorias do conteúdo referencial (não teorias do conteúdo cognitivo ou modo de apresentação). Muitos filósofos concordariam que o conteúdo referencial, que é normativo no sentido anteriormente estabelecido, não é conteúdo restrito<sup>31</sup>. Por definição, dois indivíduos que são réplicas físicas no tempo *t* “da pele para dentro” devem ter os mesmos estados de conteúdo restrito em *t*. Proponentes de teorias teleológicas não acreditam que o conteúdo referencial seja restrito. Por razões levantadas por Putnam (1975) e Burge (1979, 1986), essa perspectiva também é

---

<sup>31</sup> N.T.: *Narrow Content*.

compartilhada por outros filósofos que pensam que o conteúdo referencial é superveniente (em parte) de coisas externas aos pensadores individuais, como características de seu ambiente físico e social e/ou sua história. Em geral, os proponentes de teorias teleológicas do conteúdo mostraram pouco interesse na noção de conteúdo restrito, visto que tendem a rejeitar a afirmação de que as ciências cognitivas deveriam se limitar a usar noções restritas. Ainda assim, uma teoria teleológica do conteúdo mental poderia ser combinada com a visão de que as ciências cognitivas precisam de uma noção restrita de conteúdo. Teorias teleológicas do conteúdo tentam explicar a natureza de normas psico-semânticas (isto é, normas semânticas que o são na medida em que se aplicam a representações mentais). Em alguma medida, são questões distintas: se tais normas desempenham algum papel nas ciências cognitivas e se uma noção estrita é necessária em seu lugar ou em adição a elas.

Um ponto adicional sobre os objetivos gerais é que teorias teleológicas do conteúdo mental usualmente não tem a intenção de ser teorias sobre como entendemos significados ou nos tornamos conscientes deles. Plausivelmente, entender um significado é um estado intencional sofisticado que envolve representações de significados e não somente representações com significados. Para compreender como entendemos significados podemos recorrer a teorias psicológicas da posse de conceitos e acesso introspectivo a estruturas conceituais. Tais teorias supõem que há representações com conteúdo, enquanto teorias teleológicas do conteúdo mental tentam explicar a natureza da intencionalidade no nível mais fundamental; almejam dizer como, para início de conversa, podemos ter quaisquer representações com conteúdo.

Um ponto final sobre os objetivos gerais é que teorias teleológicas do conteúdo mental usualmente se apresentam como teorias naturais reais. Elas não tentam descrever o critério que usamos no dia a dia para identificar as crenças e desejos das pessoas, os critérios usados em atribuições intencionais na psicologia popular<sup>32</sup> (apesar de Price (2001) ser uma exceção). Aqueles que oferecem teorias naturais reais do conteúdo mental pensam que nossa habilidade cotidiana de reconhecer estados intencionais não nos faz especialistas na natureza fundamental dos estados intencionais, não mais do que nossa habilidade cotidiana de reconhecer água não nos faz especialistas na natureza fundamental da água. A ideia é que

---

<sup>32</sup> N.T: *Folk Psychology*.

podemos reconhecer instâncias de uma espécie tendo como base aparências substanciais de coisas da espécie, enquanto permanecemos ignorantes de sua natureza essencial. Desse modo, a maior parte das teorias teleológicas do conteúdo mental não acarretam que, se Bernardo pensa que Miguel sabe que hoje é terça-feira, então Bernardo deve estar pensando acerca das funções teleológicas dos sistemas de produção ou uso de representações de Miguel.

## 2. Funções Teleológicas

Como pontuado na seção anterior, uma característica crucial do conteúdo é que ele legitima avaliações semânticas. Enquanto teorias teleológicas do conteúdo mental podem vir em uma variedade de formas, elas todas partilham da ideia de que as normas que subscrevem essas avaliações dependem, ao menos parcialmente, de funções. A próxima seção explica várias ideias sobre a natureza dessa dependência. Esta seção descreve a noção de função empregada. Ela é geralmente pensada como sendo em algum sentido teleológica e normativa, mas ambos “teleológico” e “normativo” demandam qualificação. Consideraremos primeiro o primeiro termo.

Conversas sobre funções biológicas frequentemente têm um gostinho teleológico. Por exemplo, quando dizemos que é a função do coração bombear isso parece equivalente a dizer que o coração é **para** bombear sangue, ou que coração está lá **para que** sangue seja bombeado. Há um conceito próximo relacionado, a função de um artefato ou sua finalidade: por exemplo, quando dizemos que mover o cursor é a função do mouse de computador, parece que queremos dizer que é para isso que ele serve, que está lá para essa finalidade, que isso é o que os designers o projetaram para fazer. Em linhas análogas, quando biólogos dizem que bombear sangue é a função do coração, parecem querer dizer que corações foram selecionados para, adaptados para, e nesse sentido projetados para, bombear sangue. Entretanto, nesse caso, a seleção é “natural”, ou melhor, é um processo não intencional.

Aqueles que favorecem teorias teleológicas do conteúdo mental afirmam que a Mãe Natureza é intencional ou finalista (tem finalidades). No caso de Millikan (2002), não é claro se há uma oposição genuína, ou um desacordo terminológico, com o parágrafo precedente. A transição de metáfora, para metáfora morta, para uso literal de termos tais quais “design” e “propósito” é uma questão de grau, e Millikan parece usar “função” e “propósito biológico” como sinônimos. Contudo, Dennett (1988) afirma que não há fato independente da mente definitivo acerca de

significados e funções e que as funções de artefatos, as funções de sistemas biológicos e os conteúdos de pensamentos são todos dependentes de interpretação, de adotar ou uma postura de design ou uma postura intencional para com eles. Na visão de Dennett, a natureza deixa as funções e significados similarmente indeterminados.

Há alguns que preferem reservar o termo “teleológico” para contextos genuinamente finalistas no sentido mais literal de “finalidade” e se referir a funções biológicas como “teleonômicas”. Todavia, numa construção ampla do que significa para um conceito ser teleológico, um conceito pode ser considerado teleológico se é acerca do que algo é **para**, e a noção daquilo que algo foi selecionado **para** conta como teleológico nesse sentido. Esse é o sentido do termo “teleológico” usado neste texto.

Intuitivamente, o conceito de função parece ser também normativo, pois biólogos rotineiramente falam sobre sistemas funcionando normalmente ou apropriadamente, assim como funcionando mal, de modo disfuncional, danosamente e assim por diante. Aqueles que oferecem teorias teleológicas do conteúdo mental concordam que a noção relevante de função permite a possibilidade de mal funcionamento; permite que um traço característico pode ter a função de fazer Z mesmo lhe faltando a disposição para fazer Z. Por exemplo, a glândula pineal de João pode ter a função de secretar melatonina mesmo que não possa secretar melatonina porque está funcionando mal. Se é ou não apropriado descrever isso como “normativo” é mais controverso, mas entre aqueles que oferecem teorias teleológicas do conteúdo mental o desacordo é mais terminológico do que substancial, visto que usualmente tudo que se quer dizer ao dizer que a noção é normativa, é que permite a possibilidade de funcionar mal.

Alguns preferem reservar o termo “normativo” para contextos prescritivos. Nesse vocabulário, uma afirmação conta como normativa somente se implica uma afirmação de dever sem adição de outras premissas. Proponentes de teorias teleológicas do conteúdo mental podem concordar que nenhuma afirmação de dever se segue de uma atribuição de função sem a adição de outras premissas. Para uma discussão sobre isso, confira Jacob (2001) e que funções não são prescritivas. Diferentes práticas dominam diferentes discursos, mas conversas de “normas” puramente descritivas são bem estabelecidas em alguns contextos (por exemplo, normas estatísticas). Se normas psicosemânticas ou funcionais fossem prescritivas, a tentativa de naturalizá-las pareceria ignorar o aviso de Hume de cuidado ao tentar derivar afirmações de dever de afirmações de fato. Contudo, aqueles que oferecem teorias teleológicas do conteúdo mental afirmam que normas tanto de conteúdos quanto de funções recaem no lado dos fatos na divisão fato/dever.

Aqueles que favorecem teorias teleológicas do conteúdo em geral favorecem uma teoria etiológica de funções, de acordo com a qual, itens intencionais são determinados pelo seu histórico de seleção ou por seleção passada de coisas daquele tipo. Aproximadamente, em uma teoria etiológica de funções, a função de um item é aquilo para o que ele foi selecionado, ou aquilo para o que coisas daquele tipo foram selecionadas.

Wright (1973, 1976) ofereceu a primeira defesa desenvolvida de uma teoria etiológica por um filósofo, mas expressões anteriores dessa ideia podem ser encontradas em escritos de alguns biólogos (*vide* AYLA, 1970). A análise proposta por Wright é a seguinte:

- A função de X é Z se e somente se,  
1. Z é uma consequência (resultado) de X estar lá,  
2. X está lá porque faz (resulta em) Z

Wright pretendeu que essa fórmula funcionasse para uma variedade ampla de atribuição de funções; para artefatos, assim como para partes de organismos, e para atribuição de função tanto em “biologia criacionista” quanto em biologia evolutiva. Por essa razão, pretendeu que o “faz” do segundo requerimento fosse sem tempos verbais. Portanto, pretende-se que o segundo requerimento seja lido como requerendo que X está lá porque faz, fez ou fará Z estar lá. O tempo verbal específico dependeria da natureza da consequência relevante da etiologia, a história causal que explica X estar lá por causa do seu efeito Z.

Os detalhes dessa fórmula são seguidamente considerados problemáticos. Por exemplo, não é claro como ela possibilita o mau funcionamento, dado o primeiro requerimento. O segundo requerimento também é muito amplo para capturar consequências etiológicas conferidas a funções. Wright (2010) concede que, se um homem que estivesse segurando um poste causasse condução de eletricidade por si, e isso o impedisse de soltar o poste, então, em sua análise original, a função do homem ao segurar o poste seria a de permitir que a corrente elétrica passasse através de si. À luz desse problema, Wright emenda sua análise ao propor que o que é necessário é um tipo de etiologia da consequência particular, uma “etiologia virtuosa” na qual a consequência implicada na etiologia deve ser uma “virtude”. No caso de funções derivadas da seleção natural, ele vê uma consequência “virtuosa” como uma consequência adaptativa.



Outros que ofereceram teorias etiológicas da função abandonam o primeiro requerimento de Wright e falam mais especificamente da seleção como o processo de fundo responsável pela etiologia da consequência. Nessa visão, a função de um item é, aproximadamente, aquilo para o que ele é selecionado (ou itens do mesmo tipo). Millikan (1989a) e Neander (1983, 1991) argumentam que as funções da biologia podem ser concebidas de modo mais preciso por uma teoria de fundo e que sua análise não precisa ser uma análise da linguagem cotidiana. Neander (1991, p. 74) fornece a seguinte definição para funções em fisiologia:

É a/uma função própria de um item ( $X$ ) de um organismo ( $O$ ) fazer aquilo que itens de mesmo tipo que  $X$  fizeram ao contribuir para a aptidão de ancestrais de  $O$ , e que causou o genótipo do qual  $X$  é uma expressão fenotípica, selecionada por seleção natural.

Teorias etiológicas da função biológica precisam ter espaço para o fato de que traços de ancestrais podem ter sido selecionados para algo diferente da presente função dos traços nos descendentes. Por exemplo, as nadadeiras de pinguins e as asas vestigiais de emas não tem mais a função de voo, mesmo que os antebraços dos ancestrais tenham sido selecionados para o voo. Griffiths (1993) Godfrey-Smith (1994) oferecem versões a lá “história moderna” da teoria etiológica, de acordo com as quais funções são determinadas por seleção recente. Note que a seleção não cessa quando traços “fixam-se” se há contínua manutenção seletiva corrigindo novas mutações danosas quando elas surgem. Entretanto, a seleção requer alguma variação e Schwartz (1999) sugere que um suplemento de utilidade contínua é necessário, que se ativa quando a variação se faz ausente por algum tempo. Na ausência de qualquer variação, o traço retém sua função se é ainda adaptativo.

Além disso, uma questão é se a teoria etiológica é circular (ver NANAY, 2011). A preocupação é que se um traço-instância<sup>33</sup> é tipificado por sua função e se a função do traço depende do histórico de seleção relacionado ao tipo relevante, a análise é circular. Neander e Rosenberg (no prelo) respondem que a função de um traço e seu tipo em termos de função específica é co-superveniente da história de seleção e que há somente a aparência superficial de circularidade. Para desvendar

---

<sup>33</sup> N.T.: *trait token*.

se o traço *X* tem a função de *Z*, dizem eles, primeiro identifique a linhagem de traços ao qual *X* pertence; uma linhagem de traços conecta o traço ancestral e descendente por mecanismos responsáveis pela herança. Então segmente a linhagem naqueles lugares em que seleção por *Z* termina e começa. *X* tem a função de *Z* somente se houve seleção para performar *X* no segmento *X* da linhagem. Este procedimento não pressupõe conhecimento prévio da função de *X* ou conhecimento do pertencimento de *X* numa classificação de traços por funções específicas. Esta é também uma proposta alternativa para lidar com vestigialidade.

Para desempenhar um papel em explicações naturalistas do conteúdo mental, o processo relevante de seleção deve ser não intencional, mas não precisa ser seleção natural operando numa escala de tempo evolutiva. Millikan (1984) oferece uma teoria etiológica de funções na qual funções podem também resultar de seleção de memes. Papineau (1984) fala de aprendizado e Dretske (1986) invoca funções que dependem de recrutamento por condicionamento. Garson (2011) argumenta que a noção de seleção deveria ser flexibilizada para que a retenção diferencial sem replicação diferencial também conte como seleção, nesse caso seleção neuronal poderia contar como uma forma de seleção que poderia subscrever as funções que subscrevem conteúdo. Enquanto os conteúdos de representações sensorio-perceptuais podem ser determinados por funções que derivem de seleção natural operando numa escala de tempo evolutiva, o papel do aprendizado na aquisição de conceitos sugere que outros tipos de funções que derivam de outras formas de seleção podem ser necessárias para os conteúdos de conceitos aprendidos. Não há, entretanto, acordo estabelecido sobre como melhor definir de maneira ampla a classe de funções relevantes.

Apesar de teorias etiológicas dominarem a discussão de funções normativas em filosofia da biologia, a teoria etiológica não é livre de controvérsias. Alguns questionam se a teleologia pode ser naturalizada (*vide* BEDAU, 1991). Outros dão suporte a outras teorias por razões distintas. Talvez a teoria sistêmica seja a alternativa mais popular (*vide* CUMMINS, 1975). Teorias sistêmicas da função enfatizam o papel de atribuições de função nas análises de sistemas. Análises funcionais de sistemas decompõem conceitualmente atividades complexas de sistemas inteiros em atividades de suas partes contribuintes. A função de uma parte é sua contribuição à atividade complexa do sistema sob análise. Proponentes da teoria etiológica não possuem objeções à ideia que biólogos dão a análises funcionais de sistemas, mas mantém que análises sistemáticas, em si mesmas, falham em naturalizar a normatividade de funções ou não fazem isso com sucesso. Alguns

que defendem uma teoria sistêmica argumentam que biologia não precisa de uma noção naturalista de mal funcionamento (*vide* DAVIES, 2001), enquanto outros argumentam que funções anormais são estatisticamente atípicas (BOORSE, 2002; CRAVER, 2001; LEWENS, 2004). Leitores que gostariam de ler mais sobre isso e sobre outras teorias da função podem buscar diversos volumes que surgiram recentemente (ALLEN; BEKOFF; LAUDER, 1998; BULLER, 1999; ARIEW; CUMMINS; PERLMAN, 2002).

É usual notar que funções etiológicas (teleológicas) são distintas do papel causal de funções envolvido no que usualmente foi chamado “funcionalismo” em filosofia da mente. Papéis causais de funções são seguidamente definidos como um subconjunto das disposições causalmente relevantes de um traço atual, e funcionalismo é muitas vezes definido como a visão de que estados mentais são individuados ou classificados em tipos tendo como bases tais disposições (*vide* BLOCK, 1984). Se funções com papel causal são subconjuntos de disposições atualmente possuídas por traços característicos, então eles não permitem a possibilidade de mal funcionamento porque um traço não pode ter o papel causal funcional de Z e ao mesmo tempo não ter a disposição para Z.

Dito isso, a distinção entre funcionalismo e o que poderia ser classificado “teleo-funcionalismo” é de menor contraste do que pensado inicialmente. Uma razão é que formulações do funcionalismo clássico muitas vezes falam de papéis causais **característicos** ou **normais** de estados mentais. Algumas vezes explicitamente para permitir patologia (por exemplo, LEWIS, 1980). Outra razão é que, apesar de funções teleológicas seguidamente serem ditas efeitos selecionados ou efeitos a partir dos quais traços foram selecionados, tais funções também podem ser descritas como disposições selecionadas ou disposições para as quais traços foram selecionados. Ambas as formas de funcionalismo permitem múltipla realizabilidade física de traços que desempenham as mesmas funções.

### 3. Teorias Teleosemânticas

O que todas as teorias teleológicas (ou “teleosemânticas”) do conteúdo mental têm em comum é a ideia que normas psico-semânticas são, em última instância, deriváveis de normas funcionais. Além de dizer isso, é difícil dar uma definição elaborada e qualificada desse grupo de teorias.

Considere, por exemplo, algumas teorias que são claramente formuladas como alternativas à teleosemântica, como a teoria da dependência assimétrica de Fodor (1990b) ou o apelo à convergência sob condições epistêmicas ideais (vide REY, 1997). A elaboração dessas teorias está para além do escopo deste texto, mas podemos assinalar que ambas parecem precisar de uma noção de funcionamento **normal** ou **apropriado**. A teoria de Fodor aponta para o percebedor e pensador “íntacto”. Provavelmente, este é alguém cujos sistemas perceptuais e cognitivos funcionam apropriadamente (isso está de acordo com o carácter *ceterus paribus* das leis às quais a teoria de Fodor refere-se). A ideia de convergência sob condições epistêmicas ideais também envolve uma noção de funcionamento normal, visto que condições epistêmicas não são ideais se percebedores e pensadores são anormais em certos aspectos, por exemplo, se forem cegos ou psicóticos. Se o funcionamento normal ou apropriado é analisado em termos de uma teoria etiológica, que diz que sistemas funcionam normalmente somente se todas as suas partes possuem as disposições para as quais foram selecionadas, então essas teorias seriam qualificadas como teorias teleológicas do conteúdo mental sob a caracterização dada no primeiro parágrafo desta seção. Aqueles que propõem essas teorias podem rejeitar uma teoria etiológica das funções, mas eles precisam de alguma análise delas. De qualquer forma, poderia haver versões etiológicas ou teleológicas de teorias desse tipo.

Um apelo à funções teleológicas também pode ser combinado com uma variedade de outras ideias sobre como o conteúdo é determinado. Por exemplo, podem existir tanto versões isomórficas quanto informacionais da teleosemântica. No primeiro caso, a proposta pode ser que o isomorfismo relevante é um que sistemas cognitivos se adaptaram para explorar. Uma ideia alternativa é que o isomorfismo não precisa ser especificado dado que os alvos representacionais são determinados por funções teleológicas. Esta parece ser a visão de Cummins (1996, p.120), apesar de Cummins ser em geral crítico de funções teleológicas em biologia. Uma versão teleológica de uma teoria informacional é dada quando o conteúdo presumidamente depende de carregar, armazenar ou processar informações de um mecanismo. A noção relevante de informação é definida de modos variáveis, mas, grosso modo, um tipo de estado (evento, etc.) é dito portador de informação natural sobre outro estado (evento, etc.) quando é causado por ele ou corresponde a ele.

Às vezes é dito que o papel de funções em uma teoria teleológica do conteúdo é explicar como o erro é possível, em vez de explicar como o conteúdo é determinado, mas ambas andam de mãos dadas. Para observar isso podemos começar com a teoria causal do conteúdo, em sua forma bruta, e ver como o

problema do erro surge. De acordo com a teoria causal bruta, uma representação mental representa o que for que causa representações daquele tipo; *Rs* representam *Cs* se e somente se *Cs* causam *Rs*. Um problema com essa proposta simples é sua falha em fornecer possibilidades de deturpação, como apontado por Fodor (1987, 101-104). Para ver esse problema, relembre a ocasião em que um papel amassado é visto como um gato. A teoria causal bruta não permite essa caracterização do evento porque, se papel amassado causa uma instância de GATO, então papel amassado é a extensão de GATO, de acordo com a teoria causal bruta. Visto que gatos são às vezes causa de GATOs, gatos são também sua extensão. Entretanto, o problema é que papel amassado é incluído na extensão assim que causa uma instância de GATO, então de acordo com essa teoria não há espaço lógico para a possibilidade de erro, visto que candidatos a erro são transformados por sua própria ocorrência. Note que o problema é simultaneamente acrescentar as causas certas e descartar as erradas. GATO não pode ter o conteúdo **gato** a não ser que não gatos (incluindo papéis amassados) sejam excluídos do seu conteúdo. Desse modo, explicar como o conteúdo é determinado e a possibilidade do erro são tarefas complementares e não separadas.

O problema do erro é um aspecto do que (pós Fodor) é muitas vezes chamado “o problema da disjunção”. Com respeito à teoria causal bruta, o nome aplica-se porque a teoria acarreta conteúdos disjuntivos quando não deveria. Por exemplo, acarreta que GATOs, no caso considerado há pouco, têm o conteúdo **gatos ou papel amassado**. O problema da disjunção é maior que o problema do erro porque não somente em casos de erro que representações mentais são causadas por coisas que não são suas extensões (FODOR, 1990c). Suponha, por exemplo, que Miguel ao falar sobre seu cão de estimação da infância lembra Samuel do seu gato de estimação da infância. Nesse caso, nenhuma deturpação está presente, mas a teoria causal bruta novamente implica conteúdos disjuntivos inapropriados. Agora implica que GATO em Samuel tem como conteúdo algo como **gatos ou fala sobre cães de estimação**. Esse último aspecto do problema da disjunção pode ser chamado problema da representação *in absentia*: como explicamos nossa capacidade de pensar sobre coisas ausentes? Como representações mentais retêm ou obtêm seus conteúdos fora de contextos perceptuais?

Questionar sobre como alterar a teoria causal bruta para permitir erros é um lugar por onde começar a vislumbrar uma proposta mais adequada. Uma abordagem seria tentar descrever certas situações nas quais somente as causas corretas podem produzir a representação em questão e manter que o conteúdo de

uma representação é o que quer que cause a representação em tais situações. Refere-se a isso ocasionalmente como uma “teoria tipo 1”. Uma teoria tipo 1 distingue entre dois tipos de situações, aquelas nas quais somente as causas corretas podem causar uma representação e aquelas em que outras coisas também podem causá-la. Uma teoria tipo 1 pode afirmar, por exemplo, que o conteúdo de uma percepção visual é quaisquer causas que possam causá-la quando o sistema perceptual está desempenhando sua função própria. O conteúdo de representações em pensamento abstrato poderia então ser derivado de seu papel na percepção. Todavia, nem todas as teorias teleológicas do conteúdo são de tipo 1. A teoria descrita na próxima seção, pode-se argumentar, é uma variante de uma teoria tipo 1, mas algumas teorias descritas nas seções posteriores não são.

A seguinte subseção descreve algumas diferenças cruciais entre teorias teleológicas. Não é possível descrever todas as teorias existentes, mas algumas abordagens diferentes são esboçadas conjuntamente com uma breve consideração acerca de suas forças e fraquezas. Objeções gerais a teorias teleológicas são discutidas na seção 4.

### 3.1 Semântica do Indicador

Stampe (1977) foi um dos primeiros filósofos na contemporaneidade a sugerir uma teoria do conteúdo de acordo com a qual conteúdo torna-se uma questão de causas confiáveis. O livro de Dretske, *Conhecimento e o Fluxo de Informação*<sup>34</sup> (1981), também foi muito influente. A teoria de Dretske desenvolvida naquele livro não é uma teoria teleológica do conteúdo mental, mas Dretske (1986, 1988, 1991) posteriormente oferece uma versão teleo-funcional da semântica do indicador. Ele começa com uma noção de portador de informação, que chama de “indicador”, e sugere que o conteúdo de uma representação é o que ela tem a função de indicar.

Drestke (1981) provê a mais cuidadosa análise da relação de indicação, seguidamente se referindo a ela em seu trabalho posterior. Entretanto, aponta para conhecimento de fundo e, visto que conhecimento é intencional, esse aspecto é omitido em sua teoria do conteúdo, ao menos na medida em que se aplica aos tipos

---

<sup>34</sup> N.T.: *Knowledge and the Flow of Information*

mais simples de representações mentais. A análise da indicação na qual sua teoria do conteúdo depende é a seguinte: um evento do tipo  $R$  indica que um estado de coisas de tipo  $C$  é o caso se e somente se a probabilidade de  $C$ , dado que  $R$  foi instanciado, é um (assumindo que certo pano de fundo ou “condições viáveis” sejam o caso).

Apesar de a indicação ser comumente subscrita por uma regularidade causal tal que  $C$ s causam  $R$ s, Dretske nos diz que não é requerido que  $C$ s causem  $R$ s.  $C$ s e  $R$ s podem ter uma causa comum, por exemplo. Ele também nos diz que o que não é necessário uma lei tal que se  $R$  então  $C$ , apesar de não poder ser meramente coincidência. Por razões locais, poderia ser que se há um  $R$  então sempre há um  $C$ . Um dos seus exemplos é o toque de campainhas: seu toque indica que alguém está na porta. Contudo, se esquilos começam a tocar campainhas porque pessoas começam a produzi-las de nozes, o toque da campainha não mais indica que alguém está na porta.

Dretske aponta que representação não equivale à indicação. “ $R$  indica  $C$ ” implica “se  $R$ , então  $C$ ”. Mas, visto que a representação errônea é possível, deve ser o caso que “ $R$  representa  $C$ ” não implica “se  $R$ , então  $C$ ”. (De qualquer modo, “ $R$  indica  $C$ ” implica “se  $R$ , então  $C$ ” se as condições relevantes viáveis ocorrerem). Desse modo, Dretske (1986) sugere que representações perceptuais *têm a função* de indicar. O ponto de partida é esse: se algo tem a função de indicar outra coisa, então supõe-se que indica, mas, como itens nem sempre desempenham suas funções, o espaço para o erro é criado. Dretske parece depender de uma análise etiológica de funções (DRETSKE, 1995, p. 7). Ele fala de estados adquirirem uma função de indicar ao serem selecionados ou recrutados para indicação. Grosso modo, Dretske sugere que  $R$ s representam  $C$ s se e somente se  $R$ s foram recrutados para indicar  $C$ s e para causarem um movimento corpóreo  $M$ .

Dretske (1995, p. 2) diz, “a ideia fundamental é que um sistema,  $S$ , representa uma propriedade,  $F$ , se e somente se  $S$  tem a função de indicar (fornecer informação sobre) a  $F$  de um certo domínio de objetos. O modo como  $S$  desempenha sua função (quando desempenha) é ocupando diferentes estados  $s_1, s_2, \dots, s_n$  correspondentes aos diferentes valores determinados  $f_1, f_2, \dots, f_n$  de  $F$ ”. Por exemplo, parte do sistema visual pode representar a orientação de linhas numa região do campo visual. Se isso for o caso, o faz porque tem a função de transpor informação sobre a orientação de linhas naquela região e desempenha essa função (quando desempenha) ao entrar em diferentes estados quando diferentes orientações de linhas estão presentes naquela região.

Essa explicação da representação parece dar espaço ao erro porque acarreta que representações precisam indicar seus conteúdos somente durante o

recrutamento, ou num ambiente, e dadas as condições viáveis nas quais o recrutamento ocorreu; o erro torna-se possível depois daquele tempo ou em outras circunstâncias ambientais. Contudo, Dretske (1986) vê um problema com essa sugestão. Ele ilustra o problema com o caso das bactérias anaeróbicas oceânicas que possuem pequenos ímãs (magnetossomos) que, atraídos para o norte magnético, serviriam para direcionar a bactéria para uma região sedimentada relativamente livre de oxigênio no solo do oceano. Plausivelmente, a função dos magnetossomos é direcionar a bactéria às condições anaeróbicas. Se “enganamos” a bactéria ao segurar uma barra magnética próxima a ela e levarmos a bactéria a sua morte, isso parece um caso de deturpação natural. Estamos, nas palavras de Dretske, procurando pela “maneira da natureza cometer um erro”, e parece que encontramos. O problema, diz Dretske, é que é indeterminado como devemos descrever a função dos magnetossomos. Podemos plausivelmente dizer que eles têm a função de indicar sedimentos livres de oxigênio. Todavia, podemos também dizer que eles têm a função de indicação geomagnética ou até mesmo o norte magnético local. Se dissermos o último, nenhuma deturpação ocorreu. A conclusão provisória de Dretske é que não podemos contar esse caso como um caso não ambíguo de erro, dado o modo como sua teoria foi esboçada até aqui.

Uma série de problemas distintos está sob o nome “o problema da indeterminação funcional” (seção 4.1) e o exemplo do magnetossomo pode ser usado para ilustrar vários deles. Entretanto, a resposta de Dretske para o problema da indeterminação sugere que sua principal preocupação era com o que é conhecido como o problema do conteúdo distal. Este, portanto, é o seu problema: suponha que temos um sistema simples que tem somente um jeito de detectar a presença de alguma característica do ambiente. Vimos um caso como esse, pois a bactéria anaeróbica tem somente um jeito de detectar condições anaeróbicas (via campo magnético local). Nesse caso, se um estado interno indica uma característica distal (condições anaeróbicas), também indicará a característica mais proximal (norte magnético local). Além do mais, se houve seleção para indicação da característica distal, também haverá seleção para indicar a característica mais proximal (visto que é ao indicar o segundo que indica o primeiro). Dretske sugere posteriormente que, mesmo se uma criatura tem várias rotas pelas quais pode detectar uma dada característica distal (por exemplo, mesmo se a bactéria pode detectar condições anaeróbicas por meio de sensores de luz), ainda haveria uma disjunção de outras características proximais que a representação poderia contar como representativa, visto que ela poderia ainda contar como tendo a função de indicar a disjunção de



outras características proximais (ou seja, norte magnético local ou luz reduzida).

Enquanto podemos perfeitamente estar dispostos a permitir que magnetossomos em bactérias anaeróbicas não representam e nem deturpam, o problema do conteúdo distal é generalizado. Quando você vê uma cadeira do outro lado da sala, você representa como um objeto sólido 3D a uma distância de você e não como um fluxo de luz refletido dela ou como padrões de disparos da sua retina. Do contrário você não caminharia até a cadeira para se sentar. Uma teoria informacional do conteúdo deve, portanto, explicar como representações mentais representam características distais do mundo, em oposição aos itens mais proximais que carregam informação sobre essas características distais para as representações que as representam.

Dretske (1986) então modifica sua proposta e defende que uma criatura que é capaz de representar determinado conteúdo deve ser capaz de aprender um número alto de novas rotas epistêmicas para a mesma característica distal. Nesse caso, diz ele, não há disjunção fechada de estímulos mais proximais que a representação poderia ser considerada como representando. Ele fala de condicionamento nesse contexto. A representação relevante é recrutada por condicionamento para indicar a característica distal, ao invés da disjunção de características mais proximais, porque não há disjunção finita invariante no tempo de características proximais que ela tem a função de indicar. Loewer (1987) destaca que o condicionamento termina na morte, ponto no qual novas rotas epistêmicas não podem ser adquiridas. Então, na morte da criatura, haverá uma disjunção de características proximais que cada uma das representações fora recrutada para indicar. (Loewer comenta que Dretske poderia apelar para rotas epistêmicas que, em tese, poderiam ser adquiridas por uma criatura, mas está incerto se isso seria bem-sucedido).

A afirmação de que deturpações são impossíveis sem aprendizado parece problemática, uma vez que aparentemente impede representações produzidas por sistemas inatos de input, como os sistemas sensorio-perceptuais inatos. Alguns psicólogos também afirmam que alguns conceitos centrais são inatos (*vide* CAREY, 2009). Posteriormente, Dretske abandona seu requisito de condicionamento na medida em que é um requisito para posse de conteúdo, mas o mantém como um requisito para o tipo de conteúdo que pode explicar comportamentos. Para uma discussão da explicação dretskiana da eficiência causal do conteúdo, confira o ensaio de McLaughlin (1991). Isso levanta novamente a questão de como representações produzidas por analisadores inatos de inputs tem conteúdo distal.

A caracterização rígida de indicação de Dretske é considerada problemática por alguns. Uma razão é que não pode haver processos de seleção não intencionais para algo fazer  $Z$  a não ser que aquela coisa, ou coisas daquele tipo, façam  $Z$ . Corações não podem ser selecionados por seleção natural para bombear sangue a não ser que alguns corações bombeiem sangue. Similarmente, nenhum mecanismo pode ser selecionado para produzir  $R$  porque indica  $C$  a não ser que alguns  $R$ s indiquem  $C$ s. Entretanto, se algum indica, *todos* os  $R$ s devem indicar  $C$ s numa certa região do espaço-tempo, dada a caracterização rígida de indicação (visto que se  $R$ s indicam  $C$ s naquela região, então nela deve ser o caso que  $C$  é o caso, dado que uma instância  $R$ , tem probabilidade um). Portanto, onde e enquanto o recrutamento persiste,  $R$ s não podem ocorrer sem  $C$ s. Fodor (1990b) questiona se esse requerimento seria alcançado ou alcançado com frequência suficiente, dado que deturpações podem ocorrer depois. Talvez o apelo de Dretske a condições viáveis pode ajudá-lo com essa aparente dificuldade. Mesmo assim, especificar condições viáveis de modo não *ad hoc*, circular ou aludindo a fenômenos intencionais (por exemplo, de um observador não distraído) pode provar ser difícil.

Há alguns sinais nos escritos de Dretske de uma inclinação a usar uma noção menos rígida de indicação, pois ele às vezes fala do conteúdo de uma representação como o “estado maximamente indicativo”. Isso sugere que haveria estados mais minimamente indicativos, o que seria um oxímoro a partir da interpretação rígida. De qualquer modo, essa interpretação mais flexível não é desenvolvida em seus escritos e em 1981 ele oferece vários argumentos contra uma flexibilização dos requerimentos.

Um argumento a mais contra semântica do indicador envolve a afirmação de que algo se qualifica como uma representação somente se é *usado* como uma representação. Millikan (1989, p. 84-90) argumenta que o conteúdo de uma representação deve, desse modo, ser determinado pelo seu uso ou algo poderia contar como uma representação sem representar nada, o que não faz sentido. A ideia parece ser a seguinte: se o estatuto representacional e o conteúdo representacional são determinados separadamente, eles poderiam se separar, se podem se separar, algo poderia contar como uma representação por satisfazer o requerimento do estatuto representacional, sem representar nada em particular por, ao mesmo tempo, falhar em satisfazer o requerimento do conteúdo representacional.

Contudo, sempre há modos de bloquear essa conclusão. Suponha que Dretske estivesse certo sobre algo ser uma representação somente se (a) o mecanismo que a produz fosse em parte recrutado para produzi-lo porque indica

algo e também porque (b) desempenha certo papel (por exemplo,) em causar movimentos corporais. Note que (a) diz respeito à produção e também é diretamente relevante para determinação do conteúdo na teoria de Dretske. Note também que (b) diz respeito ao uso. É questionável se esses dois requerimentos são adequados para caracterizar o estatuto representacional. Mas o ponto aqui é que mesmo se Millikan estiver certa sobre o estatuto representacional ser determinado pelo uso (como é, em parte, na proposta de Dretske), não se segue que a produção de representações é irrelevante para determinar seu conteúdo (como não é na proposta de Dretske). Na proposta de Dretske, a produção de uma representação determina seu conteúdo, mas algo não conta como representação a não ser que também tenha uma função relacionada ao uso.

Millikan (2004, cap.6) também argumenta que nenhum sistema pode ter a função de produzir estados portadores de informação correlacional, mesmo se a correlação não precisasse ser cem por cento confiável. Na visão de Millikan, apesar de sistemas produtores de representações produzirem representações que carregam uma forma de informação natural quando funcionam apropriadamente, eles não têm a função de fazê-lo. (Lembre-se que mesmo que corações produzam sons de batidas quando funcionam apropriadamente, eles não têm a função de produzi-los; é um efeito colateral do funcionamento apropriado). Ela destaca que a função do seu sistema visual não pode ser garantir correlação geral entre representações de um certo tipo (por exemplo, todos os VERMELHOS produzidos por sistemas visuais humanos) e conteúdo de um certo tipo (todas as instâncias de vermelho). Seu sistema visual, por exemplo, não pode ter a função de garantir a produção de VERMELHOS somente na presença de vermelho. Se essa objeção for bem-sucedida, ela deixa em aberto a possibilidade de ser usada uma noção alternativa de informação natural, como a causal. (como discutido na seção 3.5).

Apesar de alguns problemas com a articulação detalhada da semântica do indicador de Dretske, seu *insight* central parece importante e atraente. É plausível que sistemas sensoriais-perceptuais tenham a função de produzir representações que carregam informação e que isso tenha influência em seu conteúdo. Uma tentativa alternativa de elaborar esse *insight* é esboçado posteriormente.

### 3.2 Teorias Baseadas em Benefícios e Consumidores

Millikan (1984) e Papineau (1984) foram os primeiros a oferecer versões não-informacionais baseadas em “benefícios” ou “consumidores” de teorias teleológicas do conteúdo mental. A teoria de Millikan é descrita nessa seção e a de Papineau na próxima. A visão de Millikan é detalhadamente elaborada em (1984), em (1989) provê uma versão compacta, enquanto em (2004, parte IV) elabora algo entre esses dois níveis de detalhe.

Na versão inicial, a teoria de Millikan focava bastante nos “consumidores” de representações, onde consumidores de representações são os sistemas que tem historicamente usado o mapeamento entre representações e seus conteúdos para desempenhar suas (dos consumidores) funções próprias. Já em (1989), Millikan mantém que a produção de representações mentais é irrelevante para seu conteúdo. Ela tem afirmado que atenção ao consumidor é crucial para resolver um certo problema de indeterminação funcional, uma afirmação a ser discutida na seção 4.1.

Na teoria de Millikan, quando a representação relevante é usada para comunicação entre criaturas, a produtor e o consumidor da representação são diferentes criaturas. Um dos exemplos de Millikan é de um castor batendo a cauda: o castor batendo a cauda na água é o produtor de uma representação e os consumidores são os castores ao redor que, ao terem sido avisados do perigo, mergulham em busca de abrigo. No caso de representações internas, é menos claro o que conta como produtor e consumidor. Millikan às vezes fala como se fossem diferentes subsistemas e às vezes como fossem diferentes recortes temporais do mesmo sistema, antes e depois da representação ser instanciada. Em qualquer um dos casos, um consumidor é um sistema que normalmente explora o mapeamento entre uma representação e seu representado no desempenho de sua função própria, onde “normalmente” é entendido num sentido teleológico e não em sentido estatístico.

Consumidores podem ser ou não ser sistemas cognitivos; Millikan não parece exigir que sejam sistemas cognitivos. Considere o comumente mencionado caso do sapo, que responde disparando a língua a qualquer coisa apropriadamente pequena, escura e movendo-se próximo à sua retina. Nesse caso, um consumidor relevante da representação sensorio-perceptual do sapo pode ser seu sistema digestivo. O desempenho da sua função de alimentar o sapo depende disso, e nesse sentido, explora o mapeamento entre a representação sensorio-perceptiva do sapo e seu conteúdo, que é (diz Millikan) comida de sapo.

Para descobrir o conteúdo de uma representação, diz Millikan, olhamos para as funções de seus consumidores, que são coadaptadas com os sistemas produtores. Se um sistema consumidor tem a função, então sistemas passados daquele tipo fizeram algo adaptativo que contribuiu para a preservação ou proliferação de tais sistemas na população. Sapos ancestrais tinham sistemas digestivos ancestrais, e esses fizeram coisas que contribuíram para a preservação e proliferação de tais sistemas digestivos em sapos. É a explicação dessa seleção do sistema consumidor o que mais proximamente concerne o conteúdo de uma representação. Para determinar o conteúdo de uma representação consideramos aquelas ocasiões passadas nas quais sistemas consumidores daquele tipo contribuíram para seleção daquele tipo de sistema e perguntamos qual o mapeamento entre representação e mundo era requerido para essa contribuição. De acordo com Millikan, a representação visual do sapo representa comida de sapo visto que foi somente quando havia comida de sapo onde o sapo disparou a língua que o sapo se alimentou, e foi somente então que o sistema digestivo do sapo contribuiu para a seleção de sistemas desse tipo através do uso da representação. Millikan chama aquilo que deve ser mapeado pela representação desse modo de condição Normal para o desempenho da função própria do consumidor (no modo Normal). A condição Normal é o conteúdo da representação.

Uma questão que vale a pena considerar é se uma multiplicidade de consumidores (o sistema de controle motor do sapo empregado em orientá-lo na direção do estímulo, o sistema digestivo que digere a comida, o sistema circulatório que circula os nutrientes digeridos) de uma dada representação leva a uma ambiguidade inapropriada do conteúdo. Isto dependerá se os diferentes consumidores têm diferentes condições Normais para o uso da mesma representação. Se as condições Normais para as funções dos vários sistemas que consomem a representação rotineiramente coincidem, pode-se conjecturar se as condições Normais para as funções dos sistemas produtores também coincidirão, e se sim, por qual razão precisamos focar em particular nos consumidores. Essa pode ser a razão por que, em escritos posteriores, Millikan não enfatiza funções dos consumidores sobre as funções dos produtores da mesma maneira.

Alguns argumentam que a teoria de Millikan tem vantagens em comparação com Dretske (*vide* GODFREY-SMITH, 1989; MILLIKAN, 2004). Na teoria de Millikan, uma representação,  $R$ , pode representar uma característica ambiental,  $C$ , mesmo se nunca foi inteiramente confiável que se houve um  $R$  então houve um  $C$ . É suficiente, na teoria dela, que  $R$ s mapeiem  $C$ s frequentemente o bastante para que

consumidores tenham tido os benefícios (por assim dizer) do mapeamento. Não há necessidade de prover condições do canal ou distinguir entre ambientes de recrutamento e pós-recrutamento de modo independente e especificável.

Também pode-se argumentar que Millikan resolveu o problema do conteúdo distal para conceitos inatos, assim como para conceitos aprendidos. Nem imagens retiniais, muito menos luz refletida da presa alimenta o sapo. Então, pode ser argumentado, a condição Normal para o desempenho da função própria do consumidor da representação perceptual do sapo é comida de sapo, não luz refletida da presa ou imagens retiniais. Entretanto, não é claro se a solução de Millikan para o problema do conteúdo distal sobrevive a um escrutínio mais próximo. Uma solução deve excluir itens inapropriadamente proximais, assim como incluir itens distais apropriados. Comida é incluído no conteúdo da representação perceptual do sapo, mas a questão é se itens proximais portadores de informação sobre comida para o sapo são excluídas. Comida de sapo não tem uso para o sapo se ele não pode detectá-la, e um sapo somente Normalmente detecta sua presa se luz é refletida dela de modo a gerar imagens retiniais apropriadas. Uma preocupação então surge, se a condição Normal inclui também os links mais proximais da cadeia causal.

Millikan considera uma objeção relacionada, que diz respeito a condições onipresentes benéficiais de um pano de fundo, sendo a preocupação *prima facie* se sua teoria as exclui. Para continuar no mesmo exemplo: considere que outras coisas, além de comida de sapo, foram requeridas para contribuir com a adaptabilidade<sup>35</sup> em ocasiões passadas nas quais representações perceptuais de sapos foram usadas (por exemplo, oxigênio e gravidade). A teoria dela acarreta que as representações perceptuais do sapo significam, não **comida de sapo**, mas algo como **comida de sapo na presença de oxigênio e gravidade**. Millikan exclui tais condições de fundo com base no fato de que elas não **explicam** o sucesso dos sistemas que consomem a representação.

Este texto refere-se à teoria de Millikan como “baseada em benefícios”, visto que conecta conteúdo aos benefícios para as criaturas (ou para sistemas consumidores) acumulados a partir do uso da representação. Aquilo ao que a representação se refere não é necessariamente benéfico; pode, pelo contrário, ser benéfico o evitar (evitar o perigo, como no caso da batida da cauda do castor na água). Enquanto gravidade é benéfico, ser conectado à Terra pela gravidade não

---

<sup>35</sup> N.T.: *Fitness*.

é um benefício que traz resultados a sapos devido ao uso de representações de presas. A ingestão de substâncias nutricionais, em contrapartida, é algo que resulta do uso de representações de presas. Teorias baseadas em benefícios não precisam ser teorias baseadas em consumidores, todavia, podemos falar de benefícios para sistemas produtores ou (quando a seleção relevante é seleção natural operando numa escala de tempo evolutiva), para a adaptabilidade da criatura como um todo.

Uma objeção às condições Normais de Millikan é que elas são demasiadamente específicas para conteúdos plausíveis. Considere o fato que uma variedade de circunstâncias poderia impedir uma contribuição para a adaptabilidade: uma mosca infectada ou um corvo próximo poderia significar doença ou morte, em vez de nutrição para o sapo (HALL, 1990). Foi argumentado que a teoria de Millikan possui a consequência não proposital de que a representação do sapo tem o conteúdo *comida que não está infectada, quando um corvo não está próximo, ..., etc.*

Pietroski (1992) também argumenta que a teoria de Millikan provê explicações intencionais implausíveis. Sua estória do kimu pretende enfatizar esse ponto. Os kimus são criaturas daltônicas, até uma mutação emergir, resultando em um mecanismo que produz um estado cerebral, *B*, em resposta ao vermelho. Aqueles que herdaram esse mecanismo gostam da sensação, o que os leva a subir ao topo da colina mais próxima todas as manhãs (para ver o nascer do sol ou algumas flores). O resultado é que eles evitam os predadores do amanhecer, os snorfs, que caçam no vale abaixo, e somente como resultado disso, há seleção para a mutação. Como Pietroski quer descrever o caso, *Bs* tem o conteúdo vermelho (ou *há algum vermelho*) e o kimu gosta desse estado ou busca ver coisas vermelhas. O ponto da estória é que a teoria de Millikan não permitiria que ela fosse contada desse jeito. Conforme a teoria dela, o kimu não vê o alvo visual como vermelho ou deseja ver vermelho, dado que não há mapeamento entre *Bs* e vermelhos, mas entre *Bs* e espaço livre de snorfs, o que foi crucial para a adaptabilidade do kimu (e desse modo para a seleção relevante de qualquer consumidor da representação). Na teoria de Millikan, *Bs* significam *espaços livres de snorfs* e não há representação de vermelho no cérebro de kimus.

Pietroski argumenta que bancar a consequência é radicalmente revisionista nesse caso. Testes comportamentais, diz ele, poderiam dar suporte a sua versão. Coloque uma bandeira vermelha no meio de uma multidão de snorfs e os kimus irão alegremente se juntar a eles. É consistente com sua estória que kimus contemporâneos possam nunca ter visto um snof e possam ser incapazes de reconhecer um, mesmo cara-a-cara. Intuitivamente, podemos querer afirmar que

eles não conhecem nada acerca dos snorfs. Pietrosky sugere que isso pode ser um problema para todas as teorias teleológicas do conteúdo. Contudo, é mais especificamente uma objeção às versões baseadas em benefícios (outras teorias teleológicas poderiam ter como consequência que kimu representa vermelho, ver 3.5).

Millikan concorda (2000, p.149) que sua teoria acarreta que estados B do kimu representam *menos snorfs nessa direção*. Ela argumenta que precisamos distinguir entre propriedades representadas e propriedades que causam representações. De que outro modo, pergunta ela, poderia uma tartaruga pensar *comida nessa direção*, dado que ser nutritivo é uma propriedade invisível e não poderia causar uma representação sensorio-perceptual? Deixando de lado como uma tartaruga realmente pensa, a preocupação é como uma teoria causal do conteúdo pode possibilitar representações para o que está além das características superficiais de objetos, ou como teoria causais do conteúdo podem dar conta de conceitos de tipos naturais que têm “essências” escondidas ou ocultas (por exemplo, o conceito de água, que é necessariamente composto de H<sub>2</sub>O)

Price (2001) oferece uma detalhada teoria teleológica similar a Millikan. Ela defende a interpretação de Millikan da mente de kimus, dizendo que é a que melhor explica seu comportamento. Ela apoia a ideia que a razão de se fazer atribuições de conteúdo é racionalizar comportamento, sua afirmação é que o desejo de evitar snorfs é uma razão melhor para subir ao topo da colina do que o desejo de observar o sol nascer ou ver flores vermelhas. Várias respostas são possíveis. Uma é que o desejo de assistir ao nascer do sol é razão suficiente para subir a colina. Outra é que além de explicar que eles são psicologicamente incapazes de representar corretamente a presença de snorfs quando estes estão próximos de algo vermelho, nós deixamos de ter uma explicação racional para o motivo de um kimu alegremente entrar num espaço infestado de snorfs quando lá se encontra uma bandeira vermelha. Ainda outra resposta é questionar se o papel de atribuição de conteúdos é racionalizar comportamento (como famosamente afirmado por Davidson (1985) e Dennett (1996)).

Em relação a este último ponto, alguém pode perguntar se algumas atribuições de conteúdo são cabíveis para alguns propósitos teóricos e outras para outros. Pode-se concordar que as atribuições de estados intencionais da psicologia popular servem para racionalizar comportamentos, mas questionar se esse é seu papel nas ciências cognitivas. Nesse último caso, o almejado é explicar as capacidades psicológicas humanas e (no caso da neuroetologia cognitiva) de outras criaturas. Assim sendo, uma questão a ser indagada é que atribuições de conteúdo serviriam



aos propósitos explanatórios das ciências da mente e do cérebro, ao invés das nossas intuições oriundas da psicologia popular. Neander (2006) argumenta que teorias com base em benefícios geram os conteúdos errados de acordo com teorias da percepção *mainstream* (de processamento de informação), em comparação aos casos de sistemas simples discutidos na literatura filosófica. Um princípio de tais teorias é que na visão, as propriedades invisíveis de objetos são somente representadas depois que características de superfícies visíveis são primeiramente representadas (ver PALMER, 1999). Uma preocupação é que teorias baseadas nos benefícios podem implicar que somente a propriedade invisível, mas benéfica, é representada na percepção. Além disso, Shapiro (1992) discute o papel de atribuições de conteúdo em teorias do forrageamento, o que levanta um conjunto diferente de considerações teóricas.

Millikan ocasionalmente deixa claro que sua teoria pretende ser uma versão de uma teoria do isomorfismo. De acordo com uma teoria isomórfica, representações são uma questão de espelhamento das relações entre os elementos no domínio representado e as relações entre os elementos no domínio representante. Visto que as semelhanças relevantes são relacionais, não é requerido que representações compartilhem propriedades com seus representados, a não ser propriedades abstratas relacionais. Isso faz teorias do isomorfismo mais plausíveis que teorias cruas da semelhança. Todavia, esse aspecto da teoria de Millikan não é muito desenvolvido (ver SHEA, 2012) para uma discussão do papel do isomorfismo em sua teoria).

Em grande medida a teoria de Millikan tem sido responsável pelo grande interesse, positivo e negativo, que filósofos demonstraram nessa classe geral de teorias. Seus escritos sobre o tópico são extensos e essa seção abordou somente o básico de sua perspectiva.

### 3.3 Teorias não-combinatórias

Um outro modo no qual teorias teleológicas podem divergir é com respeito aos conteúdos que almejam explicar. A teoria de David Papineau, desenvolvida ao mesmo tempo que Millikan, ajudará a ilustrar esse ponto. Papineau (1984, 1987, 1990 e 1993) desenvolve uma teoria que é *top-down*, ou não-combinatorial, na medida em que estados representacionais aos quais sua teoria mais diretamente se aplica são atitudes proposicionais plenas (desejos e crenças). Em seus primeiros escritos, Millikan as vezes parece ter uma visão similar e algumas das objeções iniciais levantadas contra ela tem como base essa interpretação (ver, por ex.,

FODOR, 1990b, p. 64-69, onde ele levanta alguns dos pontos trazidos a seguir).

Na teoria de Papineau, os conteúdos de desejos são primários e aqueles de crenças são secundários, em termos de sua derivação. De acordo com ele, a “condição real de satisfação” de um desejo é “... o efeito que é o propósito biológico do desejo produzir” (1993, 58-59), com o que ele quer dizer que “alguns mecanismos passados de seleção favoreceram aquele desejo – ou, mais precisamente, a habilidade de formar aquele tipo de desejo – em virtude daquele desejo produzir aquele efeito” (1993, 59). Assim, desejos teriam a função de nos fazer, em colaboração com nossas crenças, buscar concretizar certas condições, condições que ampliaram a adaptabilidade de pessoas no passado que tiveram esses desejos. Desejos, em geral, foram selecionados por nos fazerem concretizar condições que contribuíram à nossa adaptabilidade, e desejos particulares foram selecionados por nos fazerem buscar concretizar condições particulares. Essas condições são denominadas suas condições de satisfação e elas são os conteúdos desses desejos.

A “condição de verdade real” de uma crença, nos diz Papineau, é a condição que deve ser obtida se o desejo com o qual colabora para produzir uma ação será satisfeito pela condição concretizada por aquela ação. Um desejo que tem a função de fazer com que ingiramos comida tem como conteúdo que ingiramos comida, visto que foi selecionado por nos fazer ingerir comida, e se esse desejo colabora com a crença que nos faz ir à geladeira, o conteúdo da crença é que há comida na geladeira se nosso desejo por comida somente for satisfeito indo até lá e se é verdade que há comida na geladeira (exemplo de Papineau)

Isso parece rejeitar a hipótese da Linguagem do Pensamento, de acordo com a qual o pensamento emprega uma semântica combinatória. A linguagem é combinatória na medida em que o significado de uma sentença é uma função dos significados das palavras na sentença e suas relações sintáticas. “Rover atacou Fluff” tem um significado combinatório se seu significado é uma função dos significados de “Rover”, “Fluff” e “atacou”, juntamente de relações sintáticas (“Rover atacou Fluff” tem um significado diferente de “Fluff atacou Rover”). De acordo com alguns filósofos (especialmente FODOR, 1975), o conteúdo de atitudes proposicionais é combinatorial num sentido análogo. Por exemplo, o conteúdo de uma crença é uma função dos conteúdos dos conceitos componentes empregados na proposição acreditada, conjuntamente com suas relações sintáticas. Uma teoria teleológica do conteúdo pode ser combinatorial, pois pode manter que o conteúdo de uma representação que expressa uma proposição é determinado pelas histórias separadas das representações dos constituintes conceituais da proposição (e, talvez, pelo histórico

de seleção das regras sintáticas que se aplicam a suas relações sintáticas). A teoria de Papineau não é combinatória, ao menos para algumas atitudes proposicionais. Ao invés disso, a proposta é que os conteúdos dos conceitos são uma função do seu papel nas crenças e desejos dos quais participam.

A teoria de Papineau é baseada em benefícios e alguns dos problemas discutidos anteriormente são relevantes para uma avaliação da mesma. Por exemplo, não é claro que o que desejamos é sempre o que é benéfico à adaptabilidade. Alguém pode desejar sexo, não para reproduzir ou criar laços afetivos, mesmo que os bebês e os laços afetivos sejam cruciais para a adaptabilidade. Entretanto, essa seção não tentará fazer uma avaliação geral das forças e fraquezas dessa teoria, mas vai focar em questões peculiares de propostas não-combinatórias.

Qualquer teoria não-combinatória deve enfrentar certas objeções gerais às teorias não-combinatórias, como a de que não pode dar conta da produtividade e sistematicidade do pensamento (FODOR, 1981; 1987). Esse texto não reconstruirá esse argumento (uma reconstrução pode ser encontrada no verbete “The Language of Thought Hypothesis” da SEP), mas problemas especiais para uma versão teleológica de uma teoria não-combinatória precisam ser mencionados. Considere, por ex., o desejo de dançar no entorno de uma árvore de magnólia no brilho do céu estrelado usando duas cenouras como chifres e duas metades de um repolho como seios. Provavelmente ninguém desejou fazer isso, mas suponha que alguém desenvolve esse desejo (para contrariar Papineau), então isso é desejado pela primeira vez. De acordo com uma teoria teleológica não-combinatória, não podemos caracterizar a situação desse modo. Visto que isso nunca foi desejado antes, não há história de seleção e desse modo não há conteúdo em sua primeira ocorrência, segundo teorias desse estilo. Também é um problema para esse tipo de teoria que alguns desejos não contribuem ou não podem contribuir para sua própria satisfação (o desejo por chuva amanhã ou o desejo por imortalidade), e que alguns desejos que contribuem para sua própria satisfação não serão selecionados por isso (o desejo de fumar ou de matar as próprias crianças). Em contraste, teorias teleológicas combinatórias não tem problemas específicos com novos desejos, com desejos que não podem contribuir para a realização de sua satisfação ou desejos que tem condições de satisfação que não ampliam adaptabilidade, desde que seus conceitos constitutivos tenham histórias seletivas apropriadas ou de algum modo sejam construídos de conceitos mais simples com históricos de seleção apropriados.

Papineau pode responder concordando que algumas concessões a uma semântica combinatória devem ser feitas. Uma vez que alguns desejos e crenças

tem conteúdo, os conceitos envolvidos adquirem conteúdo a partir de seu papel neles e podem ser usados para produzir desejos novos, autodestrutivos ou causalmente impotentes. Todavia, é preciso mostrar que essa concessão não é *ad hoc*. O problema é justificar a afirmação segundo a qual o desejo de explodir um avião com um sapato explosivo é combinatório, ao passo que a crença de que há comida na geladeira não é.

### 3.4 Teorias Combinatórias mais ou menos modestas

Em contraste com a teoria de Papineau, algumas teorias teleológicas são combinatórias. De acordo com elas, uma teoria teleológica dá conta somente dos conteúdos representacionais simples e processos combinatórios são adicionalmente envolvidos em determinar o conteúdo de representações mais complexas.

Há dois tipos possíveis de processos combinatórios que podem estar envolvidos. Um opera no nível da proposição, ou no nível de representações inteiramente ao estilo de mapas ou pictoriais. Entende-se que esse tipo de processo combinatório desempenharia um papel grosseiramente análogo ao papel de uma gramática na linguagem falada, ou um papel que é grosso modo análogo aos princípios de mapeamento em cartografia ou composição pictorial na produção de imagens. Por exemplo, ele nos permitiria combinar os conceitos GATO, NO e SOFÁ para produzir o pensamento (ou desejo, ou crença) de que o gato está no sofá.

Um segundo tipo de processo combinatório que pode estar envolvido opera no nível de conceitos singulares e suas concepções associadas. Alguns pensam que conceitos mais simples poderiam ser combinados em suas concepções para formular conceitos mais sofisticados, ou para fixar a referência de conceitos mais sofisticados que continuam mais ou menos próximos dos lexemas de uma linguagem. De modo mais simples, os conceitos ADULTO, HOMEM e NÃO CASADO podem ser combinados para formar o conceito SOLTEIRO por meio de uma concepção definicional. Pode haver outras concepções envolvidas, como a concepção Wittgensteiniana de semelhança de família ou concepções de estilos protótipos.

Teorias teleológicas podem ser mais ou menos modestas em seu escopo. Uma teria modesta apenas almeja diretamente dar conta dos conteúdos representacionais simples. Dretske (1986), expressa uma visão “modesta” quando dá voz a esperança que representações mais sofisticadas podem ser construídas das representações sensório-perceptuais mais simples acomodadas por sua teoria. Contudo, ainda não

há consenso entre psicólogos ou filósofos acerca de quais representações são simples.

Uma visão modesta é que teorias teleológicas deveriam ser diretamente aplicadas apenas a representações sensório-perceptuais e motoras e a conceitos inatos (aqueles que podem ser produzidos sem aprendizado). Entretanto, até isso demanda qualificação, uma vez que é controverso quais dos nossos conceitos são inatos. Numa visão nativista radical, como a de Fodor (1981), todos ou quase todos os conceitos expressos por morfonemas lexicais (os menores componentes com significado) de uma linguagem são inatos (não aprendidos, somente ativados). Se fosse realmente assim, uma teoria que almeja dar conta dos conteúdos de todos os conceitos inatos precisaria ser bastante ambiciosa. Aqueles que propõem teorias teleológicas do conteúdo genuinamente modestas não assumem essa visão, pois afirmam que algumas representações mentais correspondentes a morfemas lexicais são sofisticadas, no sentido em que são, de algum modo, compostas a partir do ou adquiridas no uso de outras representações.

Sterelny (1990) descreve sua teoria teleológica como “modesta” porque somente tenta dar conta de representações inatas e assume que elas são um subconjunto relativamente pequeno do conjunto completo de nossas representações mentais. Para dar conta das atitudes proposicionais humanas, Sterelny mantém que uma teoria teleológica do conteúdo se deparará com “dificuldades terríveis”. Ele acredita que uma teoria teleológica para representações simples será parte de uma teoria psicosemântica completa, mas não sua totalidade. Isto contrasta com a teoria de Papineau, que se aplica mais diretamente a atitudes proposicionais. Também contrasta com a tentativa ambiciosa de Millikan (1984) de diretamente dar conta, não somente do conteúdo de todas as representações mentais, mas também dos significados de todos os proferimentos linguísticos *via* uma teoria teleológica.

Uma teoria teleológica modesta pode aproveitar-se de algumas vantagens. Obviamente, a não ser que alguns conceitos possam ser derivados de outros conceitos, teorias teleológicas teriam problemas com conceitos vazios. Por exemplo, nenhum unicórnio jamais foi indicado por UNICÓRNIO, a presença de um unicórnio nunca foi uma condição Normal para o desempenho da função própria de um consumidor de UNICÓRNIO, e o desejo de encontrar unicórnios nunca foi satisfeito de modo que as condições envolvidas na satisfação desse desejo não poderiam ter contribuído para a seleção dos mecanismos que produzem desejos desse tipo. Esse problema é evitado por uma teoria teleológica que visa diretamente tratar apenas dos conteúdos de representações simples, ao presumir que nenhuma representação simples expressa um conceito vazio (Rey (2010) questiona essa pressuposição).

É às vezes argumentado que a falta de unicórnios como condição Normal não é problemática visto que UNICÓRNIO não refere (a nada atual). Certamente, teorias não modestas entregam o conteúdo referencial correto. É uma questão se uma teoria do conteúdo referencial precisa determinar a extensão de um conceito em todos os mundos possíveis. (Se a visão do leitor é que não há unicórnios em nenhum mundo possível porque unicórnios são essencialmente ficcionais, o leitor deve aqui substituir por outro exemplo de um conceito atualmente vazio, mas potencialmente não vazio, como o conceito de flogisto ou de enteléquias). Algumas teorias do conteúdo referencial encaram essa tarefa e outras não.

O maior desafio daqueles que ofertam teorias modestas será explicar como conceitos complexos podem ser compostos ou derivados de conceitos simples. É justo dizer que não é a tarefa de uma teoria fundamental do conteúdo mental *per se* explicar como conceitos complexos podem ser compostos de conceitos mais simples, mas é um problema para teorias modestas se tal explicação não se encontra disponível. Além disso, prover tal explicação é geralmente considerado problemático. Alguns dizem que teorias “modestas” têm sérias consequências nada modestas. Alega-se, por exemplo, que deve haver, por princípio, uma distinção entre analítico e sintético. Veja, por exemplo, Fodor e Lepore (1992), que argumentam que devemos escolher entre três opções: defender uma distinção entre analítico e sintético, aceitar o holismo acerca do significado ou aceitar que virtualmente nenhum conceito mais ou menos ao nível dos lexemas de uma linguagem é composto de conceitos mais simples. Além disso, eles argumentam que as duas primeiras opções não são viáveis. Todavia, alguns psicólogos mantêm que devemos de algum modo ter nossos conceitos sofisticados “inicializados” a partir de conceitos mais simples (ver CAREY, 2009). Também é importante mencionar que alguns filósofos não foram convencidos pelos argumentos de Fodor e Lepore. (leitores que gostariam de ler mais sobre conceitos e concepções podem começar com a introdução e com os textos em Margolis e Laurence (1999) bem como os verbetes da SEP sobre conceitos e sobre a distinção analítico-sintético).

### 3.5 Teorias Causais-Informacionais

Para completar esse estudo das perspectivas, retornamos às teorias informacionais, para olhar para alguns trabalhos recentes que de modo amplo

encontram-se na tradição de Stampe e Dretske. Essas teorias seriamente consideram a ideia de que representações mentais têm funções informacionais.

Primeiro, uma resposta é oferecida para um argumento que pretende bloquear todas as versões informacionais da teleosemântica. Argumenta-se que devido ao fato de funções serem *efeitos* selecionados, qualquer apelo a funções representacionais deve ser um apelo aos efeitos das representações e não às suas causas (MILLIKAN, 1989b, p. 85; PAPINEAU, 1998, p.3). Uma resposta é aceitar a conclusão do argumento, mas mantendo que um requerimento informacional adicional pode ser adicionado a um apelo a funções; teorias teleológicas do conteúdo mental podem apelar para outras coisas que não funções (SHEA, 2007).

Uma resposta alternativa rejeita o argumento. Neander (2012) afirma que sistemas sensorio-perceptuais têm “funções de resposta”, nas quais responder a algo é ser causado por este algo de modo a fazer algo outro. Por exemplo, um sistema visual pode ser causado por uma instanciação vermelha a mudar para um estado VERMELHO, pode ter sido selecionado (ao menos parcialmente) por estar disposto a mudar para um estado VERMELHO em resposta ao vermelho, tendo a função de fazê-lo.

Na visão de Neander, essas mudanças de estado representam as causas às quais o sistema supostamente responde ao produzir a representação em questão. Eles são, por assim dizer, as causas Normais do produtor de uma representação, ao invés das condições Normais para o desempenho da função própria do consumidor de representações. Nessa perspectiva, VERMELHO tem o conteúdo *vermelho* se o sistema visual que a produz tem a função de produzi-la em resposta ao vermelho, ou mais especificamente, em resposta ao vermelho ser instanciado no campo receptivo do processamento perceptual responsável pela produção de VERMELHO. Esta é a ideia básica, embora complicações sejam adicionadas. Uma delas visa resolver o problema do conteúdo distal da seguinte maneira:

Uma representação sensorio-perceptual, *R*, num sistema sensorio-perceptual, *S*, tem o conteúdo descritivo *C* e não Prox-*C* se:

- 1 *S* foi selecionado para produzir *Rs* em resposta a *Cs*; e
- 2 Se *S* foi selecionado para produzir *Rs* em resposta a ambos *Cs* e Prox-*Cs*, foi selecionado para produzir *Rs* em resposta a Prox-*C* porque esse era um meio para produzir *Rs* em resposta a *Cs*, e não vice-versa.

O segundo requerimento pretende determinar apropriadamente o conteúdo distal e aplica-se somente depois da aplicação do primeiro requerimento. O primeiro requerimento sozinho não determina satisfatoriamente o conteúdo distal porque há uma cadeia causal de  $C$  à  $R$  e, se o sistema foi selecionado por responder a  $C$ s produzindo  $R$ s, também deve ter sido selecionado por responder aos itens proximais na cadeia causal (como a luz refletida por  $C$ s na retina do olho, no caso de percepção visual). Esses itens mais proximais na cadeia causal transmitem informação sobre  $C$  para o sistema e através do sistema para  $R$ . Contudo, há uma assimetria, para a qual o segundo requerimento apela. O sistema foi selecionado por sua disposição para responder aos itens mais proximais pois dessa maneira ele respondia aos itens mais distais, mas o sistema não foi selecionado para responder aos itens mais distais pois dessa maneira ele respondia aos itens mais proximais. (Ele não responde aos itens mais proximais ao responder aos mais distais; não é assim que se dá a análise entre meios e fins).

Nessa teoria causal, um sistema sensorio-perceptual não precisa ter produzido  $R$ s somente na presença de  $C$ s durante a seleção do sistema. Não há necessidade de especificar condições do canal ou condições nas quais a representação é confiável. Essa não é uma teoria teleológica do conteúdo mental tipo 1. A ideia de representações serem causadas de modo confiável ou correlacionadas aos seus conteúdos em certas condições não figura nessa proposta.

O primeiro requerimento garante atribuições de conteúdo diferentes daquelas geradas nas teorias baseadas em benefícios. Por exemplo, considere novamente o kimu (ver seção 3.2). Como estipulado por Pietrosky, é a presença de vermelho e não a ausência de snorf que causa o mecanismo relevante para a produção de um estado  $B$ . Mecanismos desse tipo não foram selecionados por sua disposição de, na ausência de snorfs, produzir estados  $B$ . Eles não tinham tal disposição, então não poderiam ser selecionados por isso. Os mecanismos relevantes no caso do kimu foram selecionados devido a uma disposição de produzir estados  $B$  ao serem afetados causalmente por vermelho, bem como por adicionalmente causar certos movimentos (subir uma colina pela manhã). Eles foram selecionados para isso porque vermelho correlaciona-se suficientemente bem com a ausência de snorfs, no habitat de kimus. Todavia, nessa proposta, esse fato adicional se torna um fato do pano de fundo evolutivo que não é constitutivo do conteúdo. O candidato a conteúdo *menos snorfs nessa direção* falha em passar no primeiro requerimento.



Considere também o notório caso do sapo. É plausível que os circuitos visuais relevantes no cérebro do sapo foram selecionados por sua disposição de, ao serem causalmente afetados por certa configuração de características visuais (grosso modo, algo ser pequeno, preto e se mover), produzir a representação sensório-perceptual em questão, assim como por sua disposição de iniciar orientação motora, e assim por diante. Foram selecionados (plausivelmente) por sua resposta preferencial à tal configuração de propriedades visuais porque coisas com essas propriedades no mais das vezes eram nutritivas para o sapo. No entanto, os circuitos neurais visuais no sapo não foram selecionados por uma disposição para responder ao valor nutricional de um estímulo. Isso porque o sistema visual normal de um sapo não tem sensibilidade causal para o valor nutricional do estímulo e não pode ter sido selecionado devido a uma sensibilidade causal que não possui. Assim, nessa proposta, o conteúdo visual da representação é *algo pequeno, preto e que se move* (ou algo nessa linha) e não *comida de sapo*. De acordo com Neander (2006), a configuração de características visuais é o estilo correto de conteúdo visual a se atribuir considerando os propósitos das explicações científicas das capacidades visuais de um anuro.

Essa proposta não parece gerar conteúdos demasiadamente específicos do tipo anteriormente mencionado, comparativamente às teorias baseadas em benefícios. Nessa teoria informacional, o sapo não representa o estímulo como não portanto uma doença infecciosa, mesmo se somente aquelas coisas pequenas, pretas e moventes que não portavam doenças infecciosas contribuíram para a sobrevivência do sapo quando ele comeu. Sistemas perceptuais só podem ter sido selecionados por suas disposições causais se sistemas passados desse tipo as possuíam. Visto que sistemas passados não tinham disposições para responder preferencialmente à ausência de doença infecciosa em estímulos visuais que eram pequenos, pretos e moventes, o fato de que contribuições para a adaptabilidade foram feitas somente naquelas ocasiões quando doenças infecciosas estavam ausentes é, novamente, um fato do contexto evolucionar que não é constitutivo do conteúdo de acordo com a presente proposta.

Uma preocupação possível é se há espaço suficiente para deturpações. Algumas discussões iniciais de teorias teleológicas do conteúdo assumiram que o conteúdo da representação do sapo deve ser *comida de sapo* ou *mosca* ou então deturpações seriam impossíveis. O sapo não estaria errado ao disparar sua língua para capturar algo pequeno, preto e movente que não fosse comida de sapo ou

uma mosca. Contudo, deturpações são possíveis nessa proposta. Uma representação que supostamente é produzida em resposta a algo pequeno, preto e movente, é ao invés disso produzida em resposta a algo grande e parado contaria como uma deturpação e um sapo neurologicamente danificado (por exemplo, um com dano no tálamo) de fato tentará capturar um variado tipo de coisas inapropriadas (a mão do cientista e até mesmo os próprios membros). A teoria informacional aqui apresentada também tem como consequência que os estados B dos kimus deturpam quando instanciados em resposta a algo que não é vermelho. Mais importante ainda, parece implicar que VERMELHOS humanos deturpam se instanciados por algo não vermelho, como pode acontecer com certos daltônicos, em ilusões de contraste de cor ou em condições visuais fora do comum.

Como Millikan (2012) e outros apontaram, há representações que não podem ser causadas por seus conteúdos, como AMANHÃ. Nenhum amanhã pode ter causado o pensamento sobre amanhã. Todavia, AMANHÃ não é uma representação sensório-perceptual o que faz com que essa não seja ou objeção a proposta *per se*. Assim como para outras teorias modestas, o desafio é explicar como conectar essa teoria modesta para alguns conteúdos mentais a uma teoria mais ampla que dá conta de todos os conteúdos dos nossos conceitos (ver seção 3.4).

#### **4. Problemas para a Teleosemântica**

O sobrevoos precedente em torno das teorias teleológicas do conteúdo não menciona exaustivamente todas as teorias teleológicas, mas ilustra algumas das características em comum e diferenças entre elas. Agora nos direcionamos a algumas objeções que foram levantadas contra a ideia geral de teleosemântica. Essa seção aborda as objeções que foram mais influentes, algumas das quais já foram mencionadas em seções anteriores.

##### **4.1 Indeterminação Funcional**

Há vários problemas potenciais de indeterminação. Além do problema do conteúdo distal, já discutido acima em relação às diferentes teorias que tratam dele de diferentes maneiras, há outros dois problemas de indeterminação. Um concerne o fato de que seleção natural é extensional (FODOR, 1990b) e o outro concerne o

fato de que seleção natural seleciona traços para papéis causais complexos (NEANDER, 1995). Ambos os problemas, talvez, podem ser atribuídos a Drestke (1986), apesar de Dretske não distingui-los do problema do conteúdo distal, o problema que ele parece primariamente interessado em resolver.

Fodor certa vez concebeu uma teoria teleológica do conteúdo mental (publicada anos depois, como Fodor 1990a). Entretanto, ele rapidamente repudiou a ideia e desde então tem sido um dos mais vigorosos críticos da ideia geral. Sua objeção principal inicialmente era que teorias teleológicas deixam os conteúdos indeterminados porque funções são indeterminadas. Indeterminação funcional, de acordo com Fodor (1990b), advém do fato que a seleção natural é extensional no seguinte sentido: se é adaptativo para um organismo, *O*, fazer algo, *M*, na presença de uma característica ambiental, *F*, e *F* é confiavelmente co-extensional com outra característica *G*, então é igualmente adaptativo para *O* fazer *M* na presença de *G*. Fodor argumenta que por isso teorias teleológicas não podem distinguir entre candidatos a conteúdos que são co-extensos no ambiente em que a criatura evoluiu.

O exemplo de Fodor é o sapo que dispara sua língua a qualquer coisa pequena, preta e movente, e desse modo se alimenta. De acordo com Fodor, se foi adaptativo ao sapo disparar sua língua para capturar moscas, então foi igualmente adaptativo dispara-la em direção a algo pequeno, preto e movente, presumindo por simplificação que moscas e coisas pretas, pequenas e moventes são confiavelmente co-extensivas no habitat do sapo. Ainda de acordo com Fodor, podemos igualmente dizer que a função do dispositivo é detectar moscas e que sua função é detectar coisas pequenas, pretas e moventes. Desse modo, se tentamos determinar o conteúdo da representação por referência a função do mecanismo de detecção, o conteúdo se mantém indeterminado. Podemos escolher descrever a função de variados modos, mas se o conteúdo depende de como escolhemos descrever a função, então não é um conteúdo naturalizado. Note que cada um dos candidatos *mosca*, *comida de sapo* e *coisa pequena, preta e movente* autoriza diferentes avaliações concernentes ao tópico das deturpações (representações errôneas). Se o sapo está representando o estímulo como uma mosca, por exemplo, deturpa qualquer coisa que é pequena, preta e movente que não é uma mosca. Se representa o estímulo como algo pequeno, preto e movente, não comete esse erro.

A resposta padrão a essa objeção começa destacando que a função de um traço é aquilo que os traços daquele tipo foram selecionados *para*, e que a noção de seleção *para* é uma noção causal (STERELNY, 1990; MILLIKAN, 1991). Um traço é selecionado *para* ao possuir uma certa propriedade somente se aquela

propriedade causalmente contribuiu para a seleção de traços daquele tipo (ver SOBER, 1984). O coração foi selecionado para bombear sangue e não para gerar o som de suas batidas, mesmo que ambos co-ocorressem. O coração foi selecionado para bombear sangue e não gerar som ao bater, dado que o bombear (e não as batidas) contribuiu para a adaptabilidade das criaturas ancestrais e por isso causalmente contribuiu para a seleção de corações. Funções podem, desse modo, distinguir entre duas propriedades que co-variam, desde que uma não tenha causado a seleção do traço. Esse ponto é em geral bem-aceito.

Todavia, seleção *para* não é suficiente para desambiguar conteúdo (GRIFFITHS; GOODE, 1995; NEANDER, 1995). No caso do sistema de detecção do sapo, ele responde a pontos pretos moventes e ajuda o sapo a pegar moscas e engolir algo nutritivo, *ambos* desempenham um papel causal na seleção dos sistemas produtores e consumidores das representações relevantes. Foi *por* detecção de pontos pretos moventes que o sapo se alimentou. Portanto, nem essa detecção e tampouco comer algo nutritivo são efeitos colaterais ou mera combinação de traços. Retornamos a esse tópico em breve.

De qualquer modo, Fodor (1996) continua a objetar que há um problema remanescente porque conteúdo, afirma ele, é mais detalhado do que histórias de seleção dão conta. Ele mantém que teorias teleológicas não podem discriminar conteúdo refinadamente o bastante quando há propriedades lógicas ou nomologicamente co-extensas. Ser triangular (ser uma figura plana fechada com três lados retos) e ser trilateral (ser uma figura plana fechada com três ângulos internos) são propriedades logicamente co-extensas. Ser um renato (criatura com rins) e ser um cordato (criatura com coração) são nomologicamente co-extensos (presume Fodor). De acordo com ele, não podemos distinguir entre respostas adaptativas na presença de uma versus a outra de duas tais propriedades. Podemos representar cada uma distintamente, mas de acordo com Fodor, históricos de seleção não são suficientemente detalhados para distinguir tais conteúdos.

Considere duas opções: ou os poderes causais de duas propriedades co-extensionais  $F$  e  $G$  são distintos ou não são. Pressuponha primeiro que não são. Em algumas teorias mais ou menos refinadas de individuação de propriedades, estas são individuadas por seus poderes causais, de modo que se não há diferença nos poderes causais de  $F$  e  $G$ , são a mesma propriedade de acordo com aquela teoria. Nesse modo de individualizar propriedades, uma representação que se refere a uma deve referir-se também a outra, de tal modo que não há problema aqui para uma teoria do conteúdo referencial. Nesse modo de pensar, se não há distinção de

poderes causais entre triangularidade e trilateralidade, qualquer diferença nas representações mentais TRIANGULAR e TRILATERAL deve ser uma diferença distinta. Pode ser uma diferença em veículo representacional, ou em outras palavras, uma diferença entre dois predicados que denotam a mesma propriedade. Consistente com isso, podem diferir em seus papéis cognitivos. Alternativamente, teorias modestas podem sustentar que essas duas representações são semanticamente complexas, nesse caso podem ser diferentes (até mesmo uma diferença referencial) em termos dos conceitos constituintes dos quais TRIANGULAR e TRILATERAL são compostos (uma menciona ângulos e outra não).

Agora pressuponha o contrário, que  $F$  e  $G$  têm poderes causais distintos. A maioria concordaria que esse é de fato o caso se  $X$  é a propriedade de ser uma criatura com rim e  $Y$  a propriedade de ser uma criatura com coração. Nesse caso, essa versão da objeção não sai do chão. Se os poderes causais das propriedades diferem, podem desempenhar diferentes papéis em históricos de seleção. Considere, por exemplo, a proposta que os conteúdos de representações sensorio-perceptuais são (por assim dizer) suas causas Normais. Um sistema pode possuir uma disposição de  $F$ s serem a causa que o leva a fazer  $M$ , sem ter a disposição de ser causado a fazer  $M$  por  $G$ s, se  $F$  e  $G$  tem poderes causais distintos (mesmo que sejam co-extensos). O sistema pode ser selecionado por uma disposição que tem, mas não pode ser selecionado pela disposição que não tem.

A objeção de Fodor evoluiu em uma objeção geral para qualquer explicação adaptativa e para a própria noção de seleção *para*. Tomaríamos muito espaço para seguir o desenvolvimento aqui. (ver FODOR; PIATELLI-PALMARINI, 2010 e especialmente BLOCK; KITCHER, 2010; SOBER, 2011, para uma discussão crítica efetiva).

Agora nos voltamos para o segundo problema funcional de indeterminação. Ele surge do fato que sistemas orgânicos são selecionados para papéis causais complexos, como anteriormente indicado. Por exemplo, um gene em um antílope pode ter sido selecionado porque (i) alterou o formato da hemoglobina, (ii) o que aumentou o consumo de oxigênio, (iii) o que permitiu que o antílope se movesse a terrenos elevados, (iv) o que deu a ele acesso a pastos mais ricos no verão, (v) melhorando seu status nutricional, sua imunidade a doenças, seu vigor ao evitar predação, sua atração de parceiros e (vi) suas chances de sobrevivência e reprodução (NEANDER, 1995). Para determinar a função de um traço, como a alteração do formato da hemoglobina, a teoria etiológica das funções nos recomenda perguntar “o que instâncias passadas fizeram que foi adaptativo e que causou seleção de traços daquele tipo?” Nesse caso, a resposta vai de (ii) até (vi). O formato alterado

da hemoglobina fez tudo isso, e tudo foi adaptativo, e tudo isso contribuiu para a seleção do traço (i.e., foi selecionado para *tudo* isso). Então tudo isso parece ser a função do traço. Sua função é o complexo papel causal pelo qual foi selecionado.

O problema para conteúdo pode ser visto quando consideramos mecanismos que produzem ou consomem representações. O dispositivo de detecção do sapo foi selecionado porque (a) respondeu a pequenos pontos pretos moventes e (b) isso ajudou o sapo a capturar esses pontos, (c) provendo nutrientes ao sapo e (d) contribuindo para as chances de sobrevivência e reprodução do sapo em várias formas. Desse modo, dispositivos ancestrais contribuíram para a seleção daquele tipo de dispositivo por intermédio de uma rota causal complexa na qual ambas, a configuração de estímulo visual e as propriedades nutricionais do estímulo, desempenham um papel. Note que isso não depende dessas características do ambiente serem co-extensas. Mesmo que nem todo ponto preto movente seja nutritivo e nem tudo que é nutritivo seja um ponto preto movente no habitat do sapo, esse problema dos papéis causais complexos persiste. O problema é que os sistemas responsáveis por produção e consumo de representações foram selecionados devido a complexos papéis causais nos quais um número variado de características ambientais estavam envolvidas.

Agar (1993) defende a ideia que a representação do sapo representa pontos pretos moventes de comida, um conteúdo que pretende incorporar todas as propriedades causalmente responsáveis para a seleção. Contrário ao que acaba de ser dito, Price (1998, 2001) afirma que *há* uma atribuição única e correta da função de cada traço, e ela elabora um conjunto de princípios para isolar tal atribuição. Enc (2002) concorda com Price que atribuições de função devem ser determinadas se qualquer teoria teleológica do conteúdo for ser bem-sucedida, mas levanta problemas para sua tentativa de mostrar que atribuições de função são apropriadamente determinadas.

Entretanto, teorias teleológicas do conteúdo não somente apontam na direção de funções. Considere novamente a teoria causal discutida na seção anterior. O conteúdo da representação do sapo não é indeterminada entre configuração de características visuais e algo nutritivo, visto que o sistema visual do sapo não foi selecionado para produzir a representação relevante em resposta ao valor nutricional do estímulo. O sistema visual do sapo não é causalmente sensível a presença ou ausência de nutrientes e não poderia ter sido selecionado para uma sensibilidade causal que não possui. O ponto mais geral aqui é que teorias teleológicas do conteúdo apelam para funções em certos modos e deve-se examinar a teoria particular para averiguar se a ele isola conteúdos suficientemente determinados.

Aos responder ao problema da indeterminação, Millikan (1991) pode ser pensada como se ancorando no fato, de acordo com sua teoria, que é a função própria do consumidor e não a do produtor da representação que determina seu conteúdo. Por exemplo, ao discutir o exemplo do magnetossomo de Dretske, ela diz “os mecanismos QUE USAM o ofertado pelo magnetossomos não se importam nem um pouco se os magnetos apontam para o norte magnético, norte geomagnético ou, digamos, a estrela do norte. A única das condições mencionadas por Dretske que é necessária para o usuário da função própria é que o magneto aponte na direção de menos oxigênio” (MILLIKAN, 1991, 63, maiúsculas no original). Entretanto, parece (para este autor) que a ênfase de Millikan não é colocada no lugar certo. Relembre que um dos consumidores da representação do sapo é o sistema de controle motor que orienta o sapo ao estímulo. Podemos descrever sua função como controlando o sapo de modo a orientá-lo para comida de sapo, mas também poderíamos descrevê-la como orientando-o na direção de pequenos pontos pretos moventes. Um mero apelo a consumidores parece realocar o problema sem resolvê-lo. No entanto, disso não se segue que a teoria de Millikan deixa o conteúdo indeterminado. O apelo de Millikan à condição Normal é o que cumpre a tarefa de desambiguar conteúdo.

Por fim, alguns proponentes de teorias teleológicas não pensam que conteúdo é determinado nos casos usados para ilustrar o suposto problema. Dennett (1995) defende que tal indeterminação do conteúdo não é problemática. Papineau (1997) mantém que conteúdo é indeterminado no caso de criaturas nas quais falta a estrutura psicológica crença-desejo. Se a criatura não tem tal estrutura vai depender em parte em como construímos esse requerimento. Não é claro se sapos têm uma estrutura psicológica crença-desejo, mesmo concedendo que possuem estados informacionais e motivacionais. Independentemente disso, Papineau está provavelmente correto que os estados informacionais e motivacionais não são tão distintos dos nossos, e talvez ele também esteja certo quando diz que conteúdo indeterminado nesse nível não é algo problemático. Vamos, entretanto, precisar resolver problemas relacionados de indeterminação de conteúdo para os estados mentais humanos.

## **4.2 Homem do Pântano**

Outra objeção que tem sido influente é a do Homem do Pântano. Exemplos ao estilo Homem do Pântano estão presentes há algum tempo. Boorse (1976)

imagina uma população de coelhos acidentalmente coalescendo e vindo a existir como um contraexemplo a teoria etiológica de funções de Wright. Boorse afirmava que poderíamos atribuir funções a partes dos coelhos mesmo que lhes faltasse um histórico de seleção. O Homem do Pântano em particular é uma objeção levantada por Davidson (1987) como uma objeção potencial a sua própria teoria histórica (mas não teleológica) do conteúdo. Quando um homem do pântano surge espontaneamente ele é uma réplica síncrona (naquele tempo, mas não estendidamente ao longo do tempo) de Davidson em certo ponto  $t$  do tempo. O histórico do Homem do Pântano difere radicalmente do de Davidson, ele veio a existir como um resultado de uma colisão puramente acidental de partículas elementares. De modo crucial, ele não compartilha nosso histórico evolutivo ou tem qualquer outro histórico próprio, seja evolutivo ou de desenvolvimento. Nem é criado por Deus ou uma cópia de Davidson gerada por uma máquina. A semelhança entre Davidson e o Homem do Pântano não é nada além de uma estupenda coincidência. A aparência de um design é enganadora, pois o Homem do Pântano de modo algum deriva de um processo com propósito, seja natural ou intencional. As partes componentes do Homem do Pântano não possuem funções de acordo com uma teoria etiológica de funções, desse modo, seus estados “cerebrais” não tem conteúdo de acordo com uma teoria teleológica do conteúdo mental.

Muitas pessoas acham esses resultados altamente contra-intuitivos, especialmente o resultado que o Homem do Pântano não tem nenhum estado intencional. Pressupondo o fisicismo, poderíamos substituir Davidson por sua réplica acidental e ninguém, nem mesmo família e amigos, detectaria uma diferença. Ele faria os barulhos que seriam interpretados pelos seus amigos e familiares como bem-humorados, interessantes e significativos, mas de acordo tanto com teorias teleológicas quanto com a própria teoria de Davidson, ele não teria ideias sobre filosofia, nenhuma percepção dos seus arredores e nenhuma crença ou desejo.

Há duas estratégias gerais para responder a essa objeção. Uma é tentar nos fazer abandonar a intuição de que o Homem do Pântano teria estados intencionais e a outra é argumentar que qualquer intuição remanescente não mostra que as teorias teleológicas estão equivocadas. Em ambos os casos é importante isolar a intuição relevante porque, em todas as versões, o Homem do Pântano teria muito do que Davidson tinha em  $t$ . Toda a atividade química no cérebro de Davidson quando entendia palavras, por exemplo, ocorreria no cérebro-análogo do Homem do Pântano. Certas descrições dessa atividade seriam aplicáveis a ambos igualmente: por exemplo, descrições físicas, químicas e formais. Além disso, é trivial que o



Homem do Pântano tem conteúdo em sentido restrito, se “conteúdo restrito” é definido como o algo que mais se aproxima de conteúdo “da pele para dentro” que mesmo assim não seja somente superveniente com relação aos estados físicos no tempo  $t$ . Por definição, qualquer conteúdo restrito que Davidson tinha em  $t$ , é um estado interno que o Homem do Pântano igualmente tinha, visto que são fisicamente indistinguíveis “da pele para dentro” em  $t$ . É uma consequência das teorias teleológicas, que não importa quais conteúdos restritos ele tem, falta ao Homem do Pântano conteúdo comumente entendido como normativo. A intuição em conflito com teorias teleológicas, portanto, é que os estados internos do homem do pântano, que são em sentido restrito idênticos aos de Davidson, seriam ou verdadeiros ou falsos, ou acurados ou não acurados no sentido usual.

É claro que, se os estados internos do Homem do Pântano tem conteúdos vero-valoráveis, eles não podem sempre coincidir com os valores de verdade dos estados de Davidson. Qualquer um provavelmente concordaria que, em  $t$ , o Homem do Pântano não pode lembrar de sua vida passada, visto que no máximo ele teria pseudo-memórias da vida de Davidson. Também concordaríamos que ele não pode corretamente pensar que está voltando para casa, para a sua esposa e para a sua sala de estar, visto que a casa e a esposa são de Davidson. Além disso, deveria ser lembrado que muitos acreditam que Putnam (1975) mostrou que conteúdos de conceitos de tipos naturais não são supervenientes apenas ao que “está na cabeça”. Se os casos gêmeos ao estilo de Putnam podem ser construídos para outras representações mentais e seus conteúdos (ver Burge, 1979, 1986), então a falta de histórico do Homem do Pântano pode ser um problema mesmo antes de considerarmos as complicações impostas pelas teorias teleológicas. Isso requer análises cuidadosas a respeito dos controversos problemas de determinar quais intuições sobre o Homem do Pântano podem contrariar o externalismo de teorias teleológicas.

Aqueles que tentam expelir qualquer intuição remanescente contra teorias teleológicas argumentam que a aparência de design pode ser enganadora (relembre que “design” aqui inclui o trabalho mecânico da seleção natural). Considere, por exemplo, os coelhos de Boorse. Pode ser intuitivo atribuir funções a suas partes análogas a olhos. Contudo, na natureza nada é tão intrincadamente organizado como se o desempenho de uma função deixasse de ser o resultado de um processo de design. É argumentado que hábitos de pensamento, que usualmente nos levam da aparência de design à atribuição de função, levam a atribuições falsas de funções em casos não realistas puramente hipotéticos (NEANDER, 1991). Dretske (1996) argumenta em favor do caso com outro exemplo imaginário. Uma réplica de seu

antigo Toyota Tercel, Tercel-gêmeo, surge como resultado de uma tempestade insana em seu quintal. É molécula por molécula igual ao seu antigo Tercel, exceto que o “marcador de gasolina” não gira em relação a quantidade de gasolina em seu “tanque”. Podemos estar tentados a dizer que está quebrado, mas Dretske afirma que não há razão para afirmar que não está funcionando, pois dizer isso pressupõe que foi projetado para fazer algo que não consegue realizar, sendo que não foi projetado de modo algum. Se devemos reconsiderar nossas intuições nesse caso, talvez devêssemos também rever o caso da intencionalidade atribuída ao Homem do Pântano, diz Dretske.

Podemos concordar com Dretske acerca do Tercel-gêmeo e ainda assim resistir à mudança de funções em direção a intencionalidade. O problema para teorias do conteúdo, em oposição às teorias da função, é exacerbado pela relação entre intencionalidade e consciência. Muitos filósofos consideram plausível que a consciência fenomenal de um indivíduo em um dado tempo é superveniente somente sobre as propriedades físicas internas daquele indivíduo naquele tempo. Se essa tese restrita de superveniência é verdadeira, então o Homem do Pântano terá consciência fenomenal quando ele vem a existir, assumindo que Davidson a tinha em *t*. Todavia, é difícil ver como podemos atribuir consciência fenomenal ao Homem do Pântano sem também lhe atribuir estados intencionais. Por exemplo, digamos que ele tenha uma sensação de vermelho. Presume-se que parecerá para ele que ele está vendo algo vermelho. Todavia, parecer para ele que está vendo algo vermelho é plausivelmente um estado intencional.

Aqui nos conectamos com outra questão importante que foge à alçada desse verbete, mas dois apontamentos podem ser feitos. Primeiro, alguns proponentes de teorias teleológicas do conteúdo não se preocupam com essa linha argumentativa porque eles rejeitam a visão da consciência como superveniente sobre estados restritos e mantêm teorias da consciência fenomenal que negam consciência ao Homem do Pântano. De acordo com alguns, consciência fenomênica é superveniente sobre conteúdo (não-restrito), se falta ao Homem do Pântano conteúdo, também deve faltar consciência fenomenal (ver especialmente DRETSKE, 1995).

Se, no entanto, qualquer proponente de teleosemântica aceitar a tese da superveniência restrita para consciência fenomenal, não pode negar que o Homem do Pântano poderia ter consciência fenomenal. Nesse caso, a objeção permanece e com força. Nesse caso, aparentemente há somente duas opções. Uma é afirmar que o Homem do Pântano pode ter sensações vermelhas sem parecer para ele que ele vê algo vermelho. A outra é defender que, apesar de parecer que o Homem

do Pântano vê algo vermelho, esse aparecer não é *vero-valorável* no sentido usual. Essa última opção se ajusta à ideia tradicional que o aparecer tem um estatuto epistêmico especial; se ajusta com a ideia de que não podemos nos enganar sobre como as coisas aparecem para nós, e que nesse contexto deturpações não são possíveis. Entretanto, isso não se encaixa bem com outra ideia, de que uma pessoa é, em princípio, sempre falível no que diz respeito até mesmo a como as coisas aparecem.

A segunda estratégia mais geral é argumentar que intuições sobre o caso do Homem do Pântano não podem mostrar que teorias teleológicas estão incorretas porque elas são irrelevantes. Pode ser argumentado que não são relevantes se uma teoria teleológica é oferecida como uma teoria da natureza real (MILLIKAN, 1996; NEADER, 1996). A analogia com a análise *a posteriori* da natureza da água é considerada oportuna nesse ponto. Lembre que XYZ é um líquido imaginário superficialmente indistinguível da água ( $H_2O$ ), apesar de ter uma constituição molecular distinta (abreviada como “XYZ”). Argumenta-se que podemos concordar que “água” e ÁGUA podem se referir a  $H_2O$  exclusivamente, mesmo que todos os membros da comunidade relevante classificassem XYZ como água caso encontrassem, dado sua ignorância sobre a composição química dela. Seguindo Kripke e Putnam, muitos foram persuadidos que “água” e ÁGUA podem ter tido como referência somente  $H_2O$  antes mesmo de sabermos que água é  $H_2O$ , porque havia referência a uma natureza desconhecida que explicava as propriedades superficiais por meio das quais geralmente reconhecemos instâncias do líquido. Nessa visão, era (em 1700) uma possibilidade epistêmica água não ser  $H_2O$ , mas não uma possibilidade metafísica, dado que água é de fato  $H_2O$ . Similarmente, pode ser argumentado que é somente uma possibilidade epistemológica e não uma possibilidade metafísica genuína que o Homem do Pântano pode ter intencionalidade.

Note que esse último ponto não é a afirmação que é uma mera possibilidade epistemológica que o Homem do Pântano possa existir. Ao invés disso, a afirmação crucial é que, mesmo que ele exista, permaneceria uma mera possibilidade epistemológica que ele teria intencionalidade genuína. É um paralelo com a afirmação acerca de água e XYZ. Mesmo que existisse XYZ numa Terra-gêmea e ela existisse em nosso universo, XYZ não seria água. Aparências superficiais favorecem o lado que diz que o Homem do Pântano tem intencionalidade, assim com favorecem XYZ ser água, mas pode ser que a “intencionalidade” do Homem do Pântano não é intencionalidade, do mesmo modo que XYZ revelou-se não ser água (é somente água-gêmea). Intuições acerca do Homem do Pântano, assim se diz, não podem decidir a questão sobre a correta análise da intencionalidade. Ao invés disso, a

decisão sobre a intencionalidade do Homem do Pântano deveria ser direcionada por uma teoria do conteúdo que melhor dá conta do tipo real. O que, por sua vez, deveria ser direcionado por outras considerações, como qual teoria entrega as atribuições de conteúdo corretas para nós e outras criaturas existentes.

Obviamente, no caso da intencionalidade, diferente do caso da água, a natureza ou essência não pode ser uma estrutura interna, se a teoria teleológica está correta. Numa teoria como essa, intencionalidade é dita ser um tipo histórico, então a natureza previamente oculta é histórica. Como proponentes de teorias teleológicas destacam, há uma aparente necessidade de tipos históricos em biologia (por exemplo, descendentes, homólogos e espécies). (Branddon-Mitchell e Jackson (1997) argumentaram que essa resposta em termos de “natureza real” não está disponível para proponentes de teorias teleológicas do conteúdo. Ver Papineau (2001) para uma resposta.)

O debate sobre individualismo metodológico também é relevante aqui, visto que questiona se a ciência *deveria* ter qualquer tipo histórico. Se aqueles que favorecem individualismo metodológico estão corretos, teorias teleológicas do conteúdo não nos provêm um bom modo científico de individuar estados psicológicos (FODOR, 1991). Um argumento para o individualismo metodológico envolve dizer que a ciência deveria individuar tipos com base em poderes causais. De modo breve, a ideia é que o negócio da ciência é vender explicações causais e poderes causais são o que é relevante para explicações causais, então a ciência deveria classificar itens com base em similaridades e diferenças de poderes causais. Visto que não há diferenças nos poderes causais entre os rins ou crenças de Davidson em *t* e os análogos do Homem do Pântano quando este surge, os rins de Davidson e os análogos de sua réplica devem pertencem ao mesmo tipo científico, assim como as crenças de Davidson e os estados análogos do Homem do Pântano. (Para uma discussão sobre o tópico ver HEIL; MELE, 1993)

Um problema com o individualismo metodológico é que é radicalmente revisionista, pelo menos para a biologia. Além disso, se classificarmos rins com base em seus poderes causais atuais, incluímos os rins-análogos do Homem do Pântano ao custo de excluir muitos rins reais, como os rins de pessoas em diálise. Os argumentos em favor do individualismo metodológico podem parecer plausíveis, mas não são usualmente acompanhados de qualquer tentativa de entender o papel que classificações históricas desempenham na biologia e em outros campos. Sendo esse o caso, temos razões para nos preocuparmos que o entendimento de classificações científicas que dá suporte ao individualismo metodológico é muito

simples. Ademais, devemos estar atentos ao fato que proponentes de teorias teleológicas afirmam que uma teoria histórica do conteúdo é necessária para capturar normas psico-semânticas. Talvez isto seja incorreto. Todavia, se correto, e se as ciências cognitivas precisam de noções normativas como essa, então o individualismo metodológico está equivocada. Portanto, o debate deve se direcionar a questões mais específicas sobre se conteúdo normativo envolve históricos de seleção, bem como se ciências cognitivas precisam de noção normativa de conteúdo.

### 4.3 Conceitos e capacidades sofisticadas

A objeção de maior peso às teorias teleológicas do conteúdo e a mais difícil de avaliar é que não é claro como tais teorias poderiam explicar nossas capacidades cognitivas e nossos conceitos mais sofisticados.

Até o momento, nenhuma teoria naturalista do conteúdo torna perfeitamente claro como pensamos sobre democracia, virtude, quarks ou o amanhã, e portanto não é um problema peculiar das teorias teleo-funcionais. Todavia, às vezes argumenta-se que teorias teleológicas do conteúdo têm um problema especial com respeito a isso (PEACOCKE, 1992). O pensamento é que elas podem ter alguma esperança de funcionar para os conteúdos que concernem coisas que impactam a adaptabilidade – *comida, abrigo, parceiros sexuais* etc. – mas que são, em princípio, incapazes de lidar com conteúdos que não tem tal impacto, ou não em um modo apropriadamente relacionado à seleção. Alguns conteúdos não podem ter impactado a adaptabilidade porque pertencem ao futuro ou são não-existentes. Outros não podem afetar a adaptabilidade em qualquer sentido seletivo porque, apesar de terem um impacto, seu impacto não é específico: por exemplo, quarks têm um impacto, mas porque são onipresentes no ambiente não podem se qualificar como conteúdo de uma representação em virtude de uma simples história evolutiva.

Essa objeção é de difícil avaliação por algumas razões. Uma é que há diversos tipos de conceitos sofisticados e capacidades, e dar conta de todos é uma tarefa enorme. Outra é que embora a objeção seja colocada como uma objeção a todas as teorias teleológicas, diferentes versões da teoria darão respostas diferentes a ela. Ainda outra razão é que podemos conceder que ainda estamos nos momentos iniciais com respeito ao desenvolvimento de teorias teleológicas e outras teorias naturalistas do conteúdo mental. Foi somente com o advento das ciências cognitivas no meio do século passado e a aceitação geral de uma perspectiva amplamente

fiscalista nas décadas seguintes que filósofos da mente direcionaram seus esforços para tentar dar uma teoria naturalista do conteúdo mental.

Tendo tudo em vista, a presente seção faz pouco mais que oferecer algumas observações sobre como algumas versões da teleosemântica fazem algumas incursões nessas questões. A maioria dos pontos que seguem foram abordados nas seções anteriores.

Deve ser enfatizado que aqueles que favorecem teorias teleosemânticas raramente restringem as funções relevantes àquelas que derivam da seleção natural operando numa escala de tempo evolutiva. Como observado antes, pode haver processos não-intencionais de seleção que operam no âmbito da cultura ou na escala de tempo do desenvolvimento ou vida de um indivíduo. Seleção de memes, condicionamento e algumas outras formas de aprendizado e seleção neural são consideradas tipos relevantes de seleção por alguns proponentes da teleosemântica.

Aqueles que favorecem teorias teleofuncionais modestas enfatizariam também que o atomismo conceitual é altamente controverso. Atomismo conceitual é a perspectiva que todos os conceitos de mais ou igual grau de um lexema de uma linguagem natural derivam seu conteúdo, constitutivamente falando, independentemente do conteúdo de todos os outros conceitos. Muitos psicólogos e alguns filósofos acreditam que alguns conceitos complexos são de algum modo compostos ou aprendidos a partir de conceitos mais simples. Crucialmente, negar o atomismo conceitual não compromete alguém com a visão que conceitos complexos são simplesmente definidos em termos de conceitos mais simples (uma discussão mais completa de conceitos e se concepções podem desempenhar qualquer papel em determinar referência está fora de nosso escopo).

Millikan poderia nesse contexto solicitar que notemos suas noções de funções próprias derivadas e adaptadas. O que Millikan chama “função própria direta” pertence a um mecanismo para o qual houve seleção. Os mecanismos que produzem padrões de camuflagem na superfície de um polvo têm a função própria direta de fazê-lo. Os padrões que aqueles mecanismos produzem e por meio dos quais desempenham essa função possuem o que ela chama “função própria derivada”, derivada da função do mecanismo de prover camuflagem. Além disso, um padrão produzido numa ocasião particular tem uma “função própria adaptativa derivada”, que é uma função relacional, no caso, para prover camuflagem naquele contexto particular no qual o polvo está situado. Millikan usa esses sentidos ampliados nos quais itens podem ter funções para tentar explicar os conteúdos de novas representações e representações produzidas como resultado de aprendizados.

Mecanismos de aprendizagem tem certas funções e quando as desempenham em circunstâncias particulares seus produtos podem ter funções próprias adaptativas derivadas em relação àquelas circunstâncias, tenham tais circunstâncias ocorrido durante a história da espécie ou não.

Millikan (2000) dá um tratamento ampliado a conceitos. Brevemente, sua visão é que concepções desempenham um papel em determinar as extensões dos conceitos com as quais são associadas. A teoria dela pressupõe mecanismos de aprendizagem inatos que são sintonizados para identificar substâncias de diferentes tipos de acordo com certos princípios. O tipo relevante de substância é aquele que explica o passado seletivo de sucesso do mecanismo de aprendizagem. Por exemplo, alguns mecanismos mentais podem ter sido selecionados para reconhecer faces de indivíduos de acordo com certos princípios operacionais, enquanto outros podem ter sido selecionados para reconhecer animais de diferentes espécies de acordo com outros princípios operacionais. Esses mecanismos podem adquirir o “propósito” de reconhecer algo mais específico, como a face de um indivíduo particular ou um animal de uma espécie particular, porque os mecanismos foram selecionados para reconhecer coisas naqueles domínios de acordo com certos princípios, e de acordo com eles, é agora uma face individual ou uma espécie específica que ele agora tem o “propósito” de reconhecer. A extensão de um conceito substância, diz ela, é a substância para cujo reconhecimento tal conceito foi selecionado.

Grandes questões relevantes para avaliar as diferentes teorias teleológicas do conteúdo continuam abertas. Num comentário esperançoso, muito trabalho de qualidade tem sido feito explorando as possibilidades de alcance de tais teorias, produzindo objeções de princípio interessantes e respondendo tais objeções de modo a resultar em versões mais desenvolvidas e melhor defendidas. Devemos também atentar para o fato que trabalho sério em teorias naturalistas do conteúdo tem sido feito apenas ao longo de algumas décadas, e não séculos, o que é pouquíssimo tempo na escala de tempo da Filosofia.

## Referência bibliográfica

- AGAR, N. What do Frogs Really Believe? **Australasian Journal of Philosophy**, v. 71, p. 1–12, 1993.
- ALLEN, C.; BEKOFF, M.; LAUDER, G. (Eds.). **Nature's Purposes: Analyses of Function and Design in Biology**. Cambridge, Mass: Bradford, MIT, 1998.

- ARIEW, A.; CUMMINS, R.; PERLMAN, M. (Eds.). **Functions: New Readings in the Philosophy of Biology and Psychology**. Oxford: Oxford University Press, 2002.
- AYALA, F. Teleological Explanations in Evolutionary Biology. **Philosophy of Science**, v. 37, p. 1–15, 1970.
- BEDAU, M. Can Biological Teleology be Naturalized? **Journal of Philosophy**, v. 88, p. 647–55, 1991.
- BLOCK, N. Advertisement for a Semantics for Psychology. In: FRENCH, P.; UEHLING, T.; WETTSTEIN, H. (Eds.). **Studies in the Philosophy of Mind (Midwest Studies in Philosophy)**. Minneapolis: University of Minnesota Press, 1986. v. 10.
- BLOCK, N.; KITCHER, P. Misunderstanding Darwin: Natural Selection's Secular Critics Get it Wrong. **Boston Review**, Março/Abril, p. 29–32, 2010.
- BOORSE, C. Wright on Functions. **The Philosophical Review**, v. 85, p. 70–86, 1976.
- BOORSE, C. A Rebuttal on Functions. In: ARIEW, A.; CUMMINS, R.; PERLMAN, M. (Eds.). **Functions: New Essays in the Philosophy of Biology**. Oxford: Oxford University Press, 2002. p. 63–112.
- BRADDON-MITCHELL, D.; JACKSON, F. The Teleological Theory of Content. **Australasian Journal of Philosophy**, v. 75, p. 474–89, 1997.
- BULLER, D. **Function, Selection and Design**. New York: State University of New York Press, 1999.
- BURGE, T. Individualism and the Mental. In: FRENCH, P.; UEHLING, T., Jr.; WETTSTEIN, H. (Eds.). **Contemporary Perspectives in the Philosophy of Language (Midwest Studies in Philosophy)**. Minneapolis: University of Minnesota Press, 1979.
- BURGE, T. Individualism and Psychology. **Philosophical Review**, v. 95, p. 3–45, 1986.
- CAREY, S. **The Origin of Concepts**. Oxford: New York, 2009.
- CHISHOLM, R. **Perceiving: A Philosophical Study**. Cornell: Cornell University Press, 1957.
- CRAVER, C. Role Functions, Mechanism and Hierarchy. **Philosophy of Science**, v. 68, p. 31–55, 2001.
- CUMMINS, R. Functional Analysis. **Journal of Philosophy**, v. 72, p. 741–765, 1975.
- CUMMINS, R. **Representations, Targets and Attitudes**. Cambridge, Mass: MIT Press, 1996.
- DAVIDSON, D. Knowing One's Own Mind. **Proceedings and Addresses of the American Philosophical Association**, v. 60, p. 441-458, 1987.



- DAVIES, P. **Norms of Nature: Naturalism and the Nature of Functions**. Cambridge, Mass: MIT Press, 2001.
- DAWKINS, R. **The Blind Watchmaker: Why the evidence of evolution reveals a universe without design**. New York: Norton, 1986.
- DENNETT, D. Evolution, Error and Intentionality. In: WILKS, Y.; PARTIDGE, D. (Eds.). **Sourcebook on the Foundations of Artificial Intelligence**. Albuquerque: University of New Mexico Press, 1988.
- DENNETT, D. **Darwin's Dangerous Idea**. New York: Simon & Schuster, 1995.
- DEVITT, M. **Coming to our Senses; A Naturalistic Program for Semantic Localism**. Cambridge: Cambridge University Press, 1996.
- DRETSKE, F. **Knowledge and the Flow of Information**. Cambridge, MA: MIT Press, 1981.
- DRETSKE, F. Misrepresentation. In: BOGDAN, R. (Ed.). **Belief: Form, Content and Function**. New York: Oxford University Press, 1986. p. 17–36
- DRETSKE, F. **Explaining Behavior**. Cambridge, MA: Bradford Books/MIT Press, 1988.
- DRETSKE, F. Dretske's replies. In: MCLAUGHLIN, B. (1991), p. 180–221.
- DRETSKE, F. **Naturalizing the Mind**. Cambridge, MA: MIT Press, 1995.
- DRETSKE, F. Absent Qualia. **Mind and Language**, v. 11, n. 1, p. 70–130, 1996.
- ENC, B. Indeterminacy of Function Attributions. In: ARIEW, A.; CUMMINS, R.; PERLMAN, M. (2002).
- FODOR, J. The Current Status of the Innateness Controversy. In: FODOR, J. **Representations**. Cambridge, MA: Cambridge University Press, 981.
- FODOR, J. **Psychosemantics: The Problem of Meaning in the Philosophy of Mind**. Cambridge, MA: MIT Press, 1987.
- FODOR, J. Psychosemantics, or: Where do truth conditions come from? In: LYCAN, W. (Ed.). **Mind and Cognition, a Reader**. Oxford: Basil Blackwell, 1990a. p. 312–337.
- FODOR, J. A Theory of Content. In: **A Theory of Content and Other Essays**. Cambridge, MA: MIT Press, 1990b.
- FODOR, J. Information and Representation. In: HANSON, P. (Ed.). **Information, Language and Cognition**. Vancouver: University of British Columbia Press, 1990c.
- FODOR, J. A Modal Argument for Narrow Content. **Journal of Philosophy**, v. 88, p. 5–25, 1991.
- FODOR, J. Deconstructing Dennett's Darwin. **Mind and Language**, v. 11, p. 246–262, 1996.

- FODOR, J.; LEPORE, E. **Holism: A Shopper's Guide**. Oxford: Blackwell, 1992.
- FODOR, J.; PIATELLI-PALMARINI, M. **What Darwin Got Wrong**. New York: Farrar, Straus and Giroux, 2010.
- GALLISTEL, R. **The Organization of Learning**. Cambridge, MA: Bradford: MIT, 1990.
- GARSON, J. Selected Effects Functions and Causal Role Functions in the Brain: The Case for an Etiological Approach to Neuroscience. **Philosophy & Biology**, v. 26, p. 547–565, 2011.
- GELMAN, S.; WELLMAN, H. Insides and Essences: Early Understandings of the Non-obvious. In: MARGOLIS & LAURENCE (1999), p. 613-637.
- GODFREY-SMITH, P. A Modern History Theory of Functions. **Noûs**, v. 28, n. 3, p. 344–362, 1984.
- GODFREY-SMITH, P. Misinformation. **Canadian Journal of Philosophy**, v. 19, n. 4, p. 533–550, 1989.
- GOODMAN, N. **Languages of Art**. Indianapolis: Hackett, 1976.
- GRICE, P. Meaning. **The Philosophical Review**, v. 66, p. 377–388, 1957.
- GRIFFITHS, P. Functional Analysis and Proper Functions. **British Journal for the Philosophy of Science**, v. 44, p. 409–422, 1993.
- GRIFFITHS, P.; GOODE, P. The Misuse of Sober's Selection for/Selection of Distinction. **Biology and Philosophy**, v. 10, p. 99–107, 1995.
- HALL, R. Does Representational Content Arise from Biological Function? **Philosophy of Science Association**, v. 1, p. 193–199, 1990.
- HEIL, J.; MELE, A. (Eds). **Mental Causation**. Oxford: Oxford University Press, 1993.
- JACOB, P. **What Minds Can Do: Intentionality in a Non-Intentional World**. Cambridge: Cambridge University Press, 1997.
- JACOB, P. Is Meaning Intrinsically Normative. **Proceedings, Meeting of the German Analytical Philosophy**. Allermagne: Bielefeld, 2001.
- KINGSBURY, J. et al. (Eds.). **Millikan and her Critics**. Oxford: Blackwell, 2012.
- KITCHER, P. Function and Design. In: **Midwest Studies in Philosophy**. XVIII, 1993.
- LETTVIN, J. et al. What the frog's eye tells the frog's brain. **Proceedings of the IRE**, v. 47, n. 11, p. 1940–1951, 1959.
- LEWENS, T. **Organisms and Artifacts, Design in Nature and Elsewhere**. Cambridge, MA: MIT, 2004.
- LEWIS, D. Mad Pain and Martian Pain. In: BLOCK, N. (Ed.). **Readings in the Philosophy of Psychology**. Cambridge, MA: Harvard University Press, 1980. v. 1p. 216–222.
- LOEWER, B. From Information to Intentionality. **Synthese**, v. 70, p. 287–317, 1987.

- MARGOLIS, E.; LAURENCE, S. (Eds.). In: **Concepts: Core Readings**. Cambridge, MA: MIT Press, 1999.
- MCLAUGHLIN, B. (Ed.). **Dretske and his Critics**. Cambridge, MA: Blackwell, 1991.
- MILLIKAN, R. **Language, Thought and Other Biological Categories**. Cambridge, MA: MIT Press, 1984.
- MILLIKAN, R. In Defense of Proper Functions. **Philosophy of Science**, v. 56, n. 2, p. 288–302, 1989a.
- MILLIKAN, R. Biosemantics. **Journal of Philosophy**, v. 86, p. 281–97, 1989b.
- MILLIKAN, R. Truth, Rules, Hoverflies and the Kripke-Wittgenstein Paradox. **Philosophical Review**, v. 99, p. 232–53, 1990.
- MILLIKAN, R. **Speaking Up for Darwin**. Cambridge, MA: Blackwell, 1991. p. 151–165.
- MILLIKAN, R. **White Queen Psychology and Other Essays for Alice**. Cambridge, MA: MIT Press, 1993.
- MILLIKAN, R. On Swampkinds. **Mind and Language**, v. 11, n. 1, p. 70–130, 1996.
- MILLIKAN, R. **On Clear and Confused Ideas: An Essay about Substance Concepts**. Cambridge: Cambridge University Press, 2000.
- MILLIKAN, R. **Varieties of Meaning**. Cambridge, Mass: MIT Press, 2004.
- NANAY, B. A Modal Theory of Content. **Journal of Philosophy**, v. 107, p. 412–431, 2010.
- NEANDER, K. Functions as Selected Effects. **Philosophy of Science**, v. 58, p. 168–184, 1991.
- NEANDER, K. Malfunctioning and Misrepresenting. **Philosophical Studies**, v. 79, p. 109–141, 1995.
- NEANDER, K. Swampman Meets Swampcow. **Mind and Language**, v. 11, n. 1, p. 70–130, 1996.
- NEANDER, K. Types of Traits: the importance of functional homologues. In: ARIEW, A.; CUMMINS, R.; PERLMAN, M. (2002), p. 390-415.
- NEANDER, K. Content for Cognitive Science. In: MCDONALD, G.; PAPINEAU, D. (Eds.). **Teleosemantics**. Oxford: Oxford University Press, 2006. p. 167–194.
- NEANDER, K. Toward an Informational Teleosemantics. In: KINGSBURY, J.; RYDER, D.; WILLIFORD, K. (Eds.). **Millikan and Her Critics**. Oxford: Blackwell, 2012, p. 21-40.
- NEANDER, K.; ROSENBERG, A. Solving the Circularity Problem for Functions: A Reply to Nanay. **Journal of Philosophy**, v. 109, p. 613–622, 2012.
- PALMER, S. **Vision Science: Protons to Phenomenology**. Cambridge, MA: MIT Press, 1999.

- PAPINEAU, D. Representation and Explanation. **Philosophy of Science**, v. 51, p. 550–72, 1984.
- PAPINEAU, D. **Reality and Representation**. Oxford: Basil Blackwell, 1987.
- PAPINEAU, D. Truth and Teleology. In: KNOWLES, D. (Ed.). **Explanation and its Limits**. Cambridge: Cambridge University Press, 1990. p. 21–44.
- PAPINEAU, D. Philosophical Naturalism Oxford: Blackwell, 1993.
- PAPINEAU, D. Teleosemantics and Indeterminacy. **Australasian Journal of Philosophy**, v. 76, p. 1–14, 1997.
- PAPINEAU, D. The Status of Teleosemantics, or How to Stop Worrying about Swampman. **Australasian Journal of Philosophy**, v. 79, p. 279–89, 2001.
- PAPINEAU, D. Review of Fodor and Piattelli-Palmarini's What Darwin Got Wrong. **Prospect**, v. 168, p. 83–84, 2010.
- PEACOCKE, C. **A study of concepts**. Cambridge, MA: MIT Press, 1992.
- PIETROSKI, P. Intentional and Teleological Error. **Pacific Philosophical Quarterly**, v. 73, p. 267–81, 1992.
- PRICE, C. Determinate Functions. **Noûs**, v. 32, p. 54–75, 1998.
- PRICE, C. **Functions in Mind: A Theory of Intentional Content**. Oxford: Clarendon Press, 2001.
- PRINZ, J. **Furnishing the Mind: Concepts and Their Perceptual Basis**. Cambridge, MA: Bradford, MIT, 2002.
- PUTNAM, H.; GUNDERSON, K. The Meaning of “Meaning”. In: GUNDERSON, K. **Language, Mind and Knowledge, Minnesota**. Minneapolis: University of Minnesota Press, 1975, p. 131–93. Reimpresso em PUTNAM, H. **Philosophical Papers: Mind, Language and Reality**. Cambridge: Cambridge University Press, 1975.
- REY, G. **Contemporary Philosophy of Mind**. Cambridge, MA: Blackwell, 1997.
- SCHULTE, P. How Frogs See the World: Putting Millikan's Teleosemantics to the Test. **Philosophia**, v. 40, p. 483–496, 2012.
- SCHWARTZ, P. Proper Function and Recent Selection. **Philosophy of Science**, v. 66, n. 3, 1999.
- SHAPIRO, L. Darwin and Disjunction: Foraging Theory and Univocal Assignments of Content. **Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association**, 1992, p. 469–480.
- SHEA, N. Consumers Need Information: supplementing teleosemantics with an input condition. **Philosophy and Phenomenological Research**, v. 75, n. 2, p. 404–435, 2007.

- SHEA, N. Millikan's Isomorphism Requirement. In: KINGSBURY, J.; RYDER, D.; WILLIFORD, K. (Eds.). **Millikan and Her Critics**. Oxford: Blackwell, 2012.
- SOBER, E. **The Nature of Selection**. Chicago: University of Chicago Press, 1984.
- STAMPE, D. Toward a Causal Theory of Linguistic Representation. In: FRENCH, P. A.; UEHLING, T. E.; WETTSTEIN, H. K. (Eds.). **Studies in the Philosophy of Language (Midwest Studies in Philosophy)**. Minneapolis: University of Minnesota Press, 1977, p. 81–102.
- STERELNY, K. **The Representational Theory of Mind: An Introduction**. Cambridge, MA: Blackwell, 1990.
- WRIGHT, L. Functions. **The Philosophical Review**, v. 82, p. 139–168, 1973.
- WRIGHT, L. **Teleological Explanation**. Berkeley, CA: University of California Press, 1976.

## (IV) Modularidade da Mente\*

Autor: Philip Robbins

Tradutor: Pedro Noguez (PPGFIL-UFRGS)

Revisor: César Fernando Meurer (UFABC)

O conceito de modularidade ganhou força na filosofia da psicologia a partir dos anos 1980, desde a publicação do livro seminal de Jerry Fodor, *The Modularity of Mind* (1983). Nas décadas que se seguiram desde que o termo ‘módulo’ e seus cognatos adentraram pela primeira vez o léxico das ciências cognitivas, a paisagem conceitual e teórica desta área mudou drasticamente. A esse respeito, especialmente digno de nota foi o desenvolvimento da psicologia evolutiva, cujos(as) proponentes adotam uma concepção menos rigorosa de modularidade do que aquela proposta por Fodor, e argumentam que a arquitetura da mente é mais generalizadamente modular do que Fodor alegara. Enquanto Fodor (1983, 2000) traça o limiar da modularidade em sistemas de nível relativamente básico que subjazem à percepção

---

\*ROBBINS, P. The Modularity of Mind. In: ZALTA, E. N. (ed.). Stanford Encyclopedia of Philosophy. Fall Edition. Stanford, CA: The Metaphysics Research Lab, 2017. Disponível em: <https://plato.stanford.edu/archives/fall2017/entries/modularity-mind/>. Acesso em: 10 out. 2021.

The following is the translation of the entry on “The Modularity of Mind” by Philip Robbins, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP’s archives at <https://plato.stanford.edu/archives/fall2017/entries/modularity-mind/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/modularity-mind/>. We’d like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.

e à linguagem, teóricos posteriores a Fodor, como Sperber (2002) e Carruthers (2006), argumentam que a mente é modular do início ao fim, incluindo os sistemas de alto nível responsáveis pelo raciocínio, planejamento, tomadas de decisão e similares. O conceito de modularidade também figurou em debates recentes em filosofia da ciência, epistemologia, ética e filosofia da linguagem, evidência adicional de sua utilidade como ferramenta para teorizar sobre a arquitetura da mente.

## 1. O que é um módulo mental?

Em sua clássica introdução à modularidade, Fodor (1983) lista nove características as quais, coletivamente, caracterizam o tipo de sistema que lhe interessa. Na ordem original de apresentação, são elas:

- 1 Especificidade de domínio;
- 2 Operação mandatória;
- 3 Acessibilidade central limitada;
- 4 Processamento rápido;
- 5 Encapsulamento de informação;
- 6 Resposta (*output*) “rasa”;
- 7 Arquitetura neural fixa;
- 8 Padrões característicos e específicos de quebra operacional;
- 9 Ritmo e sequenciamento ontogenéticos característicos;

Um sistema cognitivo conta como modular no sentido de Fodor se for modular “nalguma extensão interessante,” o que significa que possui a maioria dessas características, a um grau considerável (FODOR, 1983, p. 37). Trata-se de uma “maioria” sopesada, pois algumas marcas da modularidade são mais importantes do que outras. O encapsulamento de informação, por exemplo, é de certo modo essencial à modularidade, bem como anterior do ponto de vista explicativo em relação a várias das outras características da lista (FODOR, 1983, 2000).

Cada item da lista pede explicação. Para agilizar a exposição, vamos agrupar a maioria das características por tema, analisando-as assim em conjuntos, conforme Prinz (2006).

*Encapsulamento e inacessibilidade.* O encapsulamento informacional e a limitada acessibilidade central são dois lados da mesma moeda. Ambos os traços

pertencem ao carácter do fluxo de informação entre mecanismos computacionais, embora em direcções opostas. O encapsulamento envolve uma restrição no fluxo de informação que adentra o mecanismo, enquanto a inacessibilidade envolve uma restrição no fluxo extrusivo de informação.

Um sistema cognitivo é informacionalmente encapsulado à medida que, ao processar um dado conjunto de entradas (*inputs*), for incapaz de acessar informação armazenada em outro lugar; tudo de que o sistema dispõe é a informação contida nessas entradas mais qualquer informação que possa estar armazenada no interior do próprio sistema; por exemplo, em um banco de dados próprio. No caso da linguagem, por exemplo:

Um analisador para [uma linguagem] *L* contém uma gramática de *L*. O que ele faz ao realizar o seu trabalho é inferir, de certas propriedades acústicas de um exemplar, uma caracterização de algumas das causas distais desse exemplar (por exemplo, a intenção do falante de que sua enunciação seja exemplar de um certo tipo linguístico). Premissas dessa inferência podem incluir qualquer informação sobre a acústica do exemplar que os mecanismos de transdução sensoria provêm, sejam quais forem as informações sobre os tipos linguísticos em *L* que a gramática internamente representada provê, e nada mais. (FODOR, 1984, p. 245-246; itálicos no original)

De modo similar, no caso da percepção, entendida como tipo de inferência não-demonstrativa (ou seja, derrotável, ou não-monotônica) de “premissas” sensoriais a “conclusões” perceptuais, a alegação de que sistemas perceptuais são informacionalmente encapsulados é equivalente à alegação de que “os dados que podem contar na confirmação de hipóteses perceptuais incluem, em geral, consideravelmente menos que o organismo pode saber” (FODOR, 1983, p. 69). A ilustração clássica dessa propriedade vem do estudo de ilusões visuais, que tendem a persistir mesmo depois que o espectador é explicitamente informado sobre o carácter do estímulo. Na ilusão de Müller-Lyer, por exemplo, as duas linhas continuam parecendo ter comprimentos desiguais mesmo depois que alguém convence a si mesmo do contrário, por exemplo, medindo-as com uma régua (*vide* Figura 1, abaixo).



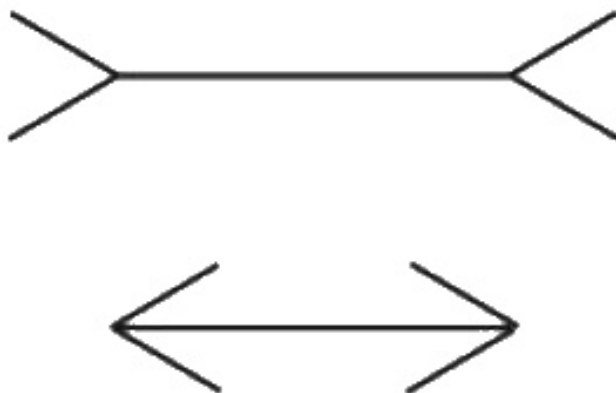


FIGURA 1. *A ilusão de Müller-Lyer.*

O encapsulamento informacional se relaciona àquilo a que Pylyshyn (1984, 1999) chama impenetrabilidade cognitiva. Mas as duas propriedades não são a mesma; antes, elas estão uma para a outra como gênero para espécie. Impenetrabilidade cognitiva é uma questão de encapsulamento relativa à informação armazenada na memória central, paradigmaticamente na forma de crenças e funções utilitárias. Mas um sistema poderia ser encapsulado sob esse aspecto sem ser encapsulado como um todo. Por exemplo, a percepção auditiva da fala poderia ser encapsulada relativamente a crenças e funções utilitárias, mas não-encapsulada com respeito à visão, como é sugerido pelo efeito McGurk (*vide* abaixo, §2.1). Do mesmo modo, um sistema poderia ser não-encapsulado relativamente a crenças e funções utilitárias, e ainda assim encapsulado relativamente à percepção; é plausível que o sistema central tenha esse caráter, na medida em que suas operações são sensíveis apenas à informação pós-perceptual, proposicionalmente codificada. Falando de maneira estrita, portanto, a impenetrabilidade cognitiva é um tipo específico de encapsulamento informacional, conquanto um tipo com especial significância arquitetural. A falta dessa característica implica falha no teste de encapsulamento, o teste decisivo da modularidade. Porém, sistemas com essa característica ainda podem falhar no teste devido à infiltração de informações de tipo diferente (isto é, não central).

O outro lado do encapsulamento informacional é a inacessibilidade ao monitoramento central. Um sistema é inacessível nesse sentido se as representações

de nível intermediário que computa antes de produzir sua resposta (*output*) são inacessíveis à consciência, e portanto indisponíveis para um relato explícito. Com efeito, sistemas centralmente inacessíveis são aqueles cujos processamentos internos são opacos à introspecção. Ainda que as respostas desses sistemas possam ser fenomenologicamente salientes, seus estados precursores não são. A compreensão linguística, por exemplo, provavelmente envolve a elaboração sucessiva de uma miríade de representações (de vários tipos: fonológica, lexical, sintática, etc.) do estímulo, embora dessas apenas o produto final, a representação do significado do que foi dito, está conscientemente disponível.

*Mandatoriedade, velocidade, e superficialidade.* Além de serem informacionalmente encapsulados e centralmente inacessíveis, sistemas modulares e processos são “rápidos, econômicos e fora de controle” (para tomar de empréstimo a expressão do roboticista Rodney Brooks). Essas características formam, como veremos, um trio natural.

A operação de um sistema cognitivo é mandatória apenas no caso em que é automática, isto é, não sob o controle consciente (BARGH; CHARTRAND, 1999). Isso significa que, queira-se ou não, as operações do sistema são ativadas pela apresentação dos estímulos relevantes, e são executadas até a conclusão do processo. Por exemplo, falantes nativas do inglês não conseguem ouvir os sons do inglês falados como mero ruído: se chegam a ouvir tais sons, elas os ouvem como inglês. Do mesmo modo, é impossível ver um arranjo de objetos em 3D no espaço como se fossem manchas coloridas em 2D, por mais que alguém se esforce.

A velocidade é, provavelmente, a marca da modularidade que menos requer explicações. Mas a velocidade é relativa, então o melhor modo de proceder aqui é por meio de exemplos. A repetição da fala (*speech shadowing*) é geralmente considerada muito rápida, com atrasos tipicamente girando em torno de 250 ms. Como a taxa silábica da fala normal é de mais ou menos 4 sílabas por segundo, isso indica que praticantes de repetição processam cada estímulo na velocidade de uma sílaba, provavelmente as menores partes identificáveis no fluxo discursivo, dado que “somente ao nível da sílaba é que nós começamos a encontrar trechos do formato de uma onda cujas propriedades acústicas chegam a estar confiavelmente relacionadas aos seus valores linguísticos” (FODOR, 1983, p. 62). Resultados similarmente impressionantes são obtidos para o caso da visão: em tarefas de apresentação sequencial rápida de estímulos visuais (conectando imagens a descrições), sujeitos experimentais obtiveram taxas de 70% de acerto com 125 ms. de exposição por imagem, e 96% de acerto com 167 ms. (FODOR, 1983, p. 63).

Em geral, um processo cognitivo conta como rápido para Fodor ele demora meio segundo, ou menos, para ocorrer.

Um traço ulterior de sistemas modulares é que as suas respostas (*outputs*) são relativamente “rasas”. Exatamente o que isso significa não é claro. Mas a profundidade de uma resposta parece ser uma função de pelo menos duas propriedades: primeiro, quão grande é a quantidade de computação necessária para produzi-la (i.e., uma resposta rasa é uma resposta computacionalmente econômica); segundo, quão constricto ou específico é o seu conteúdo informacional (i.e., uma resposta rasa é uma resposta informacionalmente geral) (FODOR, 1983, p. 87). Essas duas propriedades estão correlacionadas, na medida em que respostas cujo conteúdo é mais específico tendem a demandar mais do sistema para serem computadas, e vice-versa. Alguns autores interpretaram a baixa profundidade como requerendo um caráter não conceitual (por exemplo, CARRUTHERS, 2006, p. 4). Porém, isso entra em conflito com a conotação que Fodor dá ao termo, a qual sugere que a resposta de um sistema plausivelmente modular, tal como o do reconhecimento visual de objetos poderia ser codificada a nível de conceitos de “nível básico”, como CACHORRO e CADEIRA (ROSCH *et al.*, 1976). O impedimento aqui não incide sobre conceitos *per se*, portanto, mas antes sobre conceitos altamente teóricos como PRÓTON, que são informacionalmente muito específicos e demasiado trabalhosos do ponto de vista computacional para atender aos critérios da baixa profundidade.

Todas as três características discutidas acima, mandatoriedade, velocidade, e baixa profundidade, são associadas a, e, em alguma medida, explicáveis em termos de encapsulamento informacional. Em cada caso, menos é mais, informacionalmente falando. A mandatoriedade flui da insensitividade do sistema às funções utilitárias do organismo, o que por sua vez é uma dimensão da impenetrabilidade cognitiva. A velocidade depende da eficiência do processamento, que se correlaciona positivamente ao encapsulamento na medida em que o encapsulamento tende a reduzir a carga informacional do sistema. Em relação à baixa profundidade, a história é parecida: respostas rasas são computacionalmente econômicas, e o custo computacional é negativamente correlacionado ao encapsulamento. Em suma, quanto mais informacionalmente encapsulado for um sistema, mais provavelmente ele será rápido, econômico e fora de controle.

*Dissociabilidade e localizabilidade.* Dizer que um sistema é funcionalmente dissociável é dizer que ele pode ser seletivamente comprometido, isto é, danificado ou desabilitado com pouco ou nenhum efeito sobre a operação de outros sistemas. Como indicam os registros neuropsicológicos, comprometimentos seletivos desse

tipo tem sido frequentemente observados como uma consequência de lesões cerebrais circunscritas. Exemplos paradigmáticos do estudo da visão incluem prosopagnosia (reconhecimento facial comprometido), acromatopsia (cegueira total a cores) e acinetopsia (cegueira ao movimento); exemplos do estudo da linguagem incluem agramatismo (perda de sintaxe complexa), afasia de jargão (perda de semântica complexa), alexia (esquecimento de substantivos), e dislexia (leitura e escrita comprometidas). Cada uma dessas desordens foi encontrada em indivíduos cognitivamente normais em outros aspectos, indicando que às capacidades perdidas subjazem mecanismos funcionalmente dissociáveis.

A dissociabilidade funcional é associada à localizabilidade neural em um sentido forte. Um sistema é fortemente localizado somente no caso em que for implementado em circuitos neurais (a) relativamente circunscritos em extensão (ainda que não necessariamente em áreas contíguas) e (b) dedicados exclusivamente à operação daquele sistema. Nesse sentido, a localização vai além da mera implementação em circuitos neurais locais, visto que uma dada porção da rede neural poderia servir a mais de uma função cognitiva (ANDERSON, 2010). Candidatos que foram propostos como fortemente localizados incluem sistemas de visão cromática (V4), detecção motora (MT), reconhecimento facial (giro fusiforme), e reconhecimento de cenas espaciais (giro parahipocampal).

*Especificidade de domínio.* Um sistema é específico quanto a seu domínio à medida que é restrito o seu tópico, isto é, a classe de objetos e propriedades sobre os quais o sistema processa informação é circunscrita de maneira relativamente estrita. Como Fodor (1983, p. 103) coloca, “a especificidade de domínio tem a ver com o escopo de questões para as quais um dispositivo provê respostas (o escopo de entradas para as quais ele computa análises)”: quanto mais estreito o escopo de entradas que um sistema pode computar, tanto mais estreita é a abrangência de problemas que o sistema pode resolver – e, quanto mais estreita é a abrangência desses problemas, tão mais específico o dispositivo. Alternativamente, o grau de especificidade do domínio de um sistema pode ser entendido como uma função da gama de entradas que ativam o sistema, sendo o tamanho daquela gama de entradas o que determina o alcance informacional do sistema (CARRUTHERS, 2006; SAMUELS, 2000).

Domínios (e, por extensão, módulos) são tipicamente mais refinados do que modalidades sensoriais como visão ou audição. Isso parece claro a partir da lista

de Fodor de mecanismos cujos domínios seriam plausivelmente específicos, que inclui sistemas para percepção de cores, análise de formas visuais, análise de frases linguísticas e de reconhecimento facial e de voz (FODOR, 1983, p. 47), a nenhum dos quais correspondem faculdades perceptuais ou linguísticas em um sentido intuitivo. Também parece plausível, entretanto, que as modalidades sensoriais tradicionais (visão, audição, olfato, etc.), e a faculdade da linguagem como um todo sejam suficientemente específicas quanto a seus domínios para contar como exibindo esse traço particular da modularidade (MCCAULEY; HENRICH, 2006).

*Inatismo.* A última característica dos sistemas modulares na lista de Fodor é o inatismo, entendido como a propriedade de “desenvolver-se de acordo com padrões específicos, endogenamente determinados, sob o impacto de ativadores ambientais” (FODOR, 1983, p. 100). Deste ponto de vista, sistemas modulares são ativados mormente como resultado de processos puramente causais, como gatilhos, antes do que através de processos causais intencionais como o aprendizado. Para mais sobre essa distinção, consulte Cowie (1999); para uma análise alternativa do inatismo, baseada na noção de canalização, confira Ariew (1999). O exemplo mais familiar aqui é a linguagem, cuja aquisição ocorre em todos os indivíduos normais em todas as culturas aproximadamente no mesmo período: palavras aos 12 meses, fala telegráfica aos 18 meses, gramática complexa aos 24 meses, e assim por diante (STROMSWOLD, 1999). Outros candidatos incluem reconhecimento visual de objetos (SPELKE, 1994) e leitura mental (*mindreading*) de baixo nível (SCHOLL; LESLIE, 1999).

## **2. Modularidade à la Fodor: Uma proposta modesta**

A hipótese da modularidade modesta, como escolhemos chamar-lhe, tem duas vertentes. A primeira vertente da hipótese é positiva. Ela diz que sistemas de entrada, tais como os sistemas envolvidos na percepção e na linguagem, são modulares. A segunda vertente é negativa. Ela diz que sistemas centrais, tais como os sistemas envolvidos na fixação de crenças e raciocínios práticos, não são modulares.

Nesta seção, nós avaliamos o caso em prol da modularidade modesta. A próxima seção (§3) será dedicada à discussão da hipótese da modularidade massiva, que retém a vertente positiva da hipótese de Fodor ao mesmo tempo em que inverte a polaridade da segunda vertente, de negativa para positiva, revisando o conceito de modularidade no processo.

A parte positiva da hipótese da modularidade modesta é que sistemas de entrada são modulares. Por “sistema de entrada” Fodor (1983, p. 40) visa a um mecanismo computacional que “apresenta o mundo para o pensamento”, por meio do processamento das respostas de transdutores sensórios. Um transdutor sensório é um dispositivo que converte a energia incidente sobre as superfícies sensíveis do corpo, tais como a retina e a cóclea, em uma forma computacionalmente utilizável, sem adicionar ou subtrair informação. Grosso modo, o produto da transdução sensória é o dado sensorial bruto. O processamento da entrada envolve inferências não-demonstrativas a partir desses dados brutos até hipóteses acerca do arranjo de objetos no mundo. Essas hipóteses são então transmitidas aos sistemas centrais para propósitos de fixação de crenças, e estes sistemas, por sua vez, transmitem suas respostas a sistemas responsáveis pela produção de comportamento.

Fodor argumenta que sistemas de entrada constituem uma espécie natural, definida como “a classe dos fenômenos que contêm várias propriedades cientificamente interessantes para além das propriedades que definem a classe” (FODOR, 1983, p. 46). Ele argumenta para isso apresentando evidências de que sistemas de entrada são modulares, sendo a modularidade marcada por um conjunto de propriedades psicologicamente interessantes, a mais interessante e importante das quais sendo o encapsulamento informacional, como discutido no §1. No curso daquela discussão, nós revisamos uma amostra representativa dessas evidências, e para os presentes propósitos ela deve bastar. Leitores interessados em maiores detalhes podem consultar Fodor (1983, p. 47-101).

## **2.1 Desafios à modularidade de nível baixo**

A alegação de Fodor sobre a modularidade de sistemas de entrada foi alvo de disputa para diversos filósofos e psicólogos (CHURCHLAND, 1988; ARBIB, 1987; MARSLEN-WILSON; TYLER, 1987; MCCAULEY; HENRICH, 2006). A crítica filosófica mais abrangente é devida a Prinz (2006), que argumenta que os sistemas perceptual e linguístico raramente apresentam os traços característicos da

modularidade. Em particular, ele argumenta que esses sistemas não são informacionalmente encapsulados. Tendo isto em vista, Prinz cita dois tipos de evidência. Primeiro, parece haver efeitos intermodais na percepção, o que contaria contra o encapsulamento ao nível dos sistemas de entrada. O exemplar clássico desse tipo de efeito, também advindo da literatura sobre percepção da fala, é o efeito McGurk (MCGURK; MACDONALD, 1976). O efeito consiste em que sujeitos vendo um vídeo de um fonema sendo falado (por exemplo, / ga /) que foi sobreposto a uma gravação de som de um fonema diferente (/ ba /) ouvem um terceiro fonema totalmente diferente (/ da /). Segundo, ele aponta para o que parecem ser efeitos descendentes (*top-down effects*) sobre os processamentos visual e linguístico, cuja existência iria de encontro à hipótese da impenetrabilidade cognitiva, i.e., do encapsulamento relativo a sistemas centrais. Alguns dos casos exemplares mais notáveis de tais efeitos advêm da pesquisa sobre a percepção da fala. Provavelmente o mais conhecido é o efeito da restauração fonêmica, como no caso em que os ouvintes “preenchem” um fonema faltante em uma frase falada (*Os governadores estaduais se encontraram com seus respectivos legi\*ladores reunidos na capital*) da qual o fonema faltante (o som /s/ em *legisladores*) foi deletado e substituído pelo som de um tossido (WARREN, 1970). Por hipótese, esse preenchimento é ditado pelo entendimento dos sujeitos experimentais do contexto linguístico.

O quão convincente alguém acha essa parte da crítica de Prinz depende, contudo, do quão convincente é a explicação dele para esses mesmos efeitos. O efeito McGurk, por exemplo, parece consistente com a alegação de que a percepção da fala é um sistema informacionalmente encapsulado, embora um sistema de caráter multimodal (*vide* FODOR, 1983, p.132, n.13). Se a percepção da fala for um sistema multimodal, então o fato de que suas operações se desenrolam a partir de informações auditivas bem como visuais não necessariamente se contrapõe à alegação de que a percepção da fala é encapsulada. Outros efeitos intermodais resistem, contudo, a esse tipo de explicação. Na ilusão do *flash* duplo, por exemplo, pessoas submetidas a um único *flash* de luz acompanhado de dois *bipes* reportam terem visto dois *flashes* (SHAMS *et al.*, 2000). A mesma ideia é válida para a ilusão da mão de borracha, experimento no qual o roçar síncrono de escovas em uma das mãos do sujeito, ocluída de sua vista, e em uma mão de borracha de aparência realista que é vista por ele na posição usual da mão real escondida provoca a impressão de que a mão falsa é real (BOTVINICK; COHEN, 1998). Com respeito a fenômenos desse tipo, diferente do que ocorre com o efeito McGurk, não há candidato plausível a sistema individual, de domínio específico, cujas operações

se desenrolem a partir de múltiplas fontes de informação sensorial.

No que concerne à restauração fonêmica, poderia ser que o efeito se decorra da utilização pelos ouvintes de informação armazenada em uma base de dados propriamente linguística (especificamente, informação sobre os tipos linguísticos no léxico do português), e não em informações contextuais de alto nível. Sendo assim, não fica claro se o fenômeno da restauração fonêmica descrito acima conta como efeito descendente. Porém, nem todo caso de restauração fonêmica pode ser tão prontamente acomodado, já que o fenômeno ocorre também quando há múltiplos elementos lexicais à disposição para realizar o preenchimento (WARREN; WARREN, 1970). Por exemplo, os ouvintes preenchem as lacunas nas sentenças *A \*oda está no eixo* e *A \*oda da laranjeira está pronta* de maneiras diferentes, com um som de /r/ e com um som de /p/, respectivamente, indicando que a percepção da fala é, afinal, sensível à informação contextual.

Um outro desafio à modularidade modesta, que não foi abordado por Prinz (2006), vem de evidências de que a suscetibilidade à ilusão de Müller-Lyer varia, tanto de acordo com a cultura quanto de acordo com a idade. Parece, por exemplo, que adultos em culturas ocidentais são mais suscetíveis à ilusão do que adultos não-ocidentais; que adultos em algumas culturas não-ocidentais, como caçadores-coletores do deserto Kalahari, são praticamente imunes à ilusão; e parece que no interior de culturas ocidentais e não-ocidentais (em geral, porém nem sempre distributivamente), pré-adolescentes são mais suscetíveis à ilusão do que adultos (SEGALL; CAMPBELL; HERSKOVITS, 1966). McCawley e Henrich (2006) interpretam esses achados como indicativos de que o sistema visual é diacronicamente (e não sincronicamente) penetrável, na medida em que o modo como alguém experiencia o estímulo indutor da ilusão altera-se como resultado de sua ampla experiência perceptual ao longo do tempo. Eles também argumentam que as supramencionadas evidências de variabilidade cultural e etária na percepção militam contra a ideia de que a visão é uma capacidade inata, isto é, a ideia de que a visão está entre as “características endógenas do sistema cognitivo humano que são, senão amplamente fixadas desde o nascimento, então, ao menos, geneticamente pré-programadas” e “ativadas, mais do que formadas pela experiência subsequente do nascituro” (p. 83). Todavia, eles também apontam a seguinte ressalva:

[N]ada acerca de quaisquer dos achados que nós discutimos demonstra a penetrabilidade *sincrônica* dos estímulos de Müller-Lyer. Tampouco as descobertas



de Segall et al. (1966) providenciam evidências de que os sistemas visuais de entrada de *adultos* sejam *diacronicamente* penetráveis. Aquelas evidências sugerem que é apenas durante um estágio crucial do desenvolvimento que a suscetibilidade de seres humanos à ilusão de Müller-Lyer apresenta variação considerável e que é de variáveis *culturais* que essa variação substancialmente depende. (MCCAULEY; HENRICH, 2006, p. 99; itálicos no original)

Como tal, a evidência citada pode ser acomodada por defensores da modularidade modesta, desde que estes abram espaço para o potencial impacto de variáveis ambientais, inclusive culturais, no desenvolvimento, espaço que efetivamente é aberto pela maior parte das abordagens inatistas.

Um bom modo de mostrar isso é evocando a ideia de Segal (1996) de modularidade diacrônica (*vide* SCHOLL; LESLIE, 1999). Módulos diacrônicos são sistemas que exibem variação paramétrica ao longo de seu desenvolvimento. Por exemplo, no caso da linguagem, indivíduos diferentes aprendem a falar diferentes línguas a depender do ambiente linguístico em que crescem e, ainda assim, compartilham a mesma competência linguística subjacente, em virtude de seu (plausivelmente inato) conhecimento da Gramática Universal. Dada a observada variação na maneira como as pessoas veem a ilusão de Müller-Lyer, é possível que o sistema visual seja similarmente modular, tendo seu desenvolvimento limitado por traços do ambiente visual. Tal possibilidade parece ser consistente com a alegação de que sistemas de entrada são modulares no sentido de Fodor.

Outra fonte de dificuldades para proponentes da modularidade a nível de entrada é a evidência neurocientífica contrária à alegação de que os sistemas perceptual e linguístico são fortemente localizados. Recorde-se que, para que um sistema seja fortemente localizado, sua operação deve se realizar em sistemas neurais especialmente dedicados. A localização forte a nível de sistemas de entrada implica, então, a existência de uma correlação um a um entre sistemas de entrada e estruturas cerebrais. Como argumenta Anderson (2010, 2014), no entanto, essa correlação não existe, dado que a maior parte das regiões corticais de qualquer tamanho desempenham diferentes tarefas sobre diferentes domínios. Por exemplo, a ativação da área da face fusiforme, que outrora se pensou ser dedicada ao reconhecimento facial, é também recrutada para reconhecimento de carros e

pássaros (GAUTHIER *et al.*, 2000). De modo similar, a área de broca, que se considerava dedicada à produção da fala, também desempenha um papel no reconhecimento e no ordenamento de ações e na imagética motora (TETTAMANTI; WENIGER, 2006). Estudos de neuroimagens funcionais costumam indicar que sistemas cognitivos são, quando muito, apenas fracamente localizados, ou seja, eles são implementados em redes distribuídas do cérebro as quais se intersectam, ao invés de em regiões discretas e disjuntas.

Pode-se argumentar que o desafio mais sério à modularidade a nível de sistemas de entrada advenha, porém, das evidências de que a visão é cognitivamente penetrável, e conseqüentemente, que ela não é informacionalmente encapsulada. O conceito de penetrabilidade cognitiva, originalmente introduzido por Pylyshyn (1984), tem sido caracterizado de maneiras diversas e não-equivalentes (STOKES, 2013), mas a ideia central é esta: Um sistema perceptual é cognitivamente penetrável se e somente se suas operações são direta e causalmente sensíveis às crenças, aos desejos, às intenções ou a outros estados não-perceptuais do agente. Estudos de comportamento que se propõem a mostrar que a visão é cognitivamente penetrável datam da aurora da *new look psychology* (BRUNER; GOODMAN, 1947) e continuam sendo feitos até os nossos dias, com interesse renovado no tópico emergindo nos anos iniciais da década de 2000 (FIRESTONE; SCHOLL, 2016). Ao que parece, por exemplo, a visão é influenciada pelos estados motivacionais do agente, pois sujeitos experimentais relatam que os objetos que lhes são desejáveis parecem mais próximos (BALCETIS; DUNNING, 2010) e que figuras ambíguas aparecem conformes à interpretação associada ao resultado mais recompensador (BALCETIS; DUNNING, 2006). Ademais, a visão parece ser influenciada pelas crenças dos sujeitos, pois a categorização racial afeta relatos acerca dos tons de pele facial percebidos mesmo nos casos em que os estímulos são de igual luminescência (LEVIN; BANAJI, 2006), e a categorização de objetos afeta relatos acerca das cores percebidas de imagens em escala de cinza desses objetos (HANSEN *et al.*, 2006).

Céticos sobre a penetrabilidade cognitiva salientam, contudo, que as evidências experimentais em prol da existência de efeitos descendentes podem ser explicadas em termos de efeitos de juízo, memória e de formas relativamente periféricas de atenção (FIRESTONE; SCHOLL, 2016; MACHERY, 2015). Considere-se, por exemplo, a alegação de que arremessar uma esfera pesada (ao invés de uma esfera leve) até um alvo faz com que o alvo pareça mais distante, cuja evidência consiste nas estimativas visuais dos sujeitos sobre a distância até o alvo (WITT; PROFFITT; EPSTEIN, 2004). Se, por um lado, é possível que o maior esforço

envolvido em arremessar a esfera pesada tenha causado a aparência de maior distanciamento do alvo, por outro lado, também é possível que a estimativa de um maior distanciamento refletiu o fato de que os sujeitos destinados a arremessar a esfera pesada tenham julgado que o alvo estava a uma maior distância por acharem mais difícil de acertá-lo (FIRESTONE; SCHOLL, 2016). De fato, relatos de sujeitos de um estudo subsequente, que foram explicitamente instruídos a fazer suas estimativas tão somente com base nas aparências visuais deixaram de salientar o efeito do esforço, indicando que tal efeito era pós-perceptual (WOODS; PHILBECK; DANOFF, 2009). Outros dos aventados efeitos descendentes sobre a percepção, como o efeito da performance no jogo de golfe sobre estimativas acerca do tamanho e da distância dos buracos (WITT et al., 2008), podem ser explicados como efeitos da atenção espacial, tais como o fato de que objetos de atenção visual tendem a parecer maiores e mais próximos (FIRESTONE; SCHOLL, 2016). Estas e outras considerações relacionadas sugerem que o caso em prol da penetrabilidade cognitiva – e, por extensão, o caso contra a modularidade de baixo nível – é mais fraco do que seus proponentes dão a entender que seja.

## **2.2. O argumento de Fodor contra a modularidade de alto nível**

Volto-me agora para o lado escuro da hipótese de Fodor: a alegação de que os sistemas centrais não são modulares.

Entre os principais trabalhos dos sistemas centrais está a fixação de crenças, inclusive crenças perceptuais, via inferência não-demonstrativa. Fodor (1983) argumenta que esse tipo de processo não pode ser realizado em um sistema informacionalmente encapsulado, e portanto que sistemas centrais não podem ser modulares. Pouco melhor explanado, seu raciocínio é o seguinte:

- 1 Sistemas centrais são responsáveis pela fixação de crenças.
- 2 A fixação de crenças é isotrópica e quineana.
- 3 Processos isotrópicos e quineanos não podem ser levados a cabo por sistemas informacionalmente encapsulados.
- 4 O processo de fixação de crenças não pode ser levado a cabo por um sistema informacionalmente

- encapsulado. [a partir de 2 e 3]
- 5 Sistemas modulares são informacionalmente encapsulados.
  - 6 O processo de fixação de crenças não é modular. [a partir de 4 e 5]
- Logo:
- 7 Sistemas centrais não são modulares. [a partir de 1 e 6]

Esse argumento contém dois termos que pedem explicação, ambos relacionados à noção de holismo confirmacional na filosofia da ciência. O termo 'isotrópico' refere-se à interconexão epistêmica de crenças no sentido de que "tudo que uma cientista sabe é, em princípio, relevante para determinar em que mais ela deve acreditar. Em princípio, nossa botânica constrange nossa astronomia, se ao menos conseguirmos pensar em maneiras de conectá-las" (FODOR, 1983, P. 105). Antony (2003) apresenta um caso impactante desse tipo de amplo cruzamento interdisciplinar nas ciências, entre astronomia e arqueologia; Carruthers (2006, p. 356-357) fornece outro exemplo, ligando a física do sistema solar à teoria evolucionista. Na visão de Fodor, dado que a confirmação científica é similar à fixação de crenças, o fato de a confirmação científica ser isotrópica indica que, em geral, a fixação de crenças também tem essa propriedade.

Uma segunda dimensão do holismo confirmacional é que a confirmação é 'quineana', querendo dizer que:

[O] grau de confirmação atribuído a qualquer dada hipótese é sensível a propriedades de todo o sistema de crenças ... simplicidade, plausibilidade e conservadorismo são propriedades que teorias têm em virtude da sua relação com a inteira estrutura de crenças científicas *coletivamente tomadas*. Uma medida de conservadorismo ou de simplicidade seria uma métrica para propriedades *globais* de sistemas de crenças. (FODOR, 1983, p. 107-108; itálicos no original).

Novamente, a analogia entre o pensamento científico e o pensamento em geral subjaz à suposição de que a fixação de crenças é quineana.

A isotropia e o caráter quineano são ambas características que impossibilitam o encapsulamento, uma vez que sua posse por um sistema exigiria acesso extensivo aos conteúdos da memória central e, portanto, um alto grau de penetrabilidade cognitiva. Em termos pouco diferentes: processos isotrópicos e quineanos são 'globais' em vez de 'locais' e, uma vez que a globalidade impede o encapsulamento, ser isotrópico e ser quineano implica não ser encapsulado.

À luz de Fodor, o mote desse argumento, nomeadamente, o caráter não modular dos sistemas centrais, é uma má notícia para o estudo científico de funções cognitivas superiores. Isso é precisamente expresso pela sua "primeira lei da não-existência das ciências cognitivas", de acordo com a qual "[q]uanto mais global (por exemplo, quanto mais isotrópico) um processo cognitivo for, menos compreensível ele é" (FODOR, 1983, p. 107). A base de seu pessimismo a esse respeito é bipartite. Em primeiro lugar, sistemas globais provavelmente não são associáveis a estruturas cerebrais locais, o que os torna objetos de estudo neurocientífico nada promissores:

Nós temos visto que sistemas isotrópicos provavelmente não exibem uma arquitetura neural articulada. Se, como parece ser o caso, a arquitetura neural costuma ser um fator concomitante a restrições sobre o fluxo de informação, então a equipotencialidade neural é o que se esperaria em sistemas nos quais todo processo tem acesso mais ou menos irrestrito a todos os dados disponíveis. A moral é que, na medida em que a existência de uma correspondência entre forma e função é uma pré-condição à investigação neurofisiológica bem-sucedida, não há muito a se esperar no caminho de uma neuropsicologia do pensamento (FODOR, 1983, p. 127).

Em segundo lugar, e mais importante, processos globais são resistentes à explicação computacional, o que os converte em objetos nada promissores de estudo psicológico:

O fato é que – considerações acerca de sua realização neural à parte – sistemas globais são por si mesmos maus domínios para modelos computacionais, ao menos do tipo que cientistas cognitivos estão

acostumados a empregar. A condição para o sucesso da ciência (na física, a propósito, assim como na psicologia) é que a natureza tenha juntas nas quais se possa trinchá-la: subsistemas relativamente simples que podem ser artificialmente isolados e que se comportam, isoladamente, de maneira semelhante àquela em que se comportariam *in situ*. Módulos satisfazem essa condição; Sistemas quineanos/isotrópicos-holistas por definição não a satisfazem. Se, como eu supus, os processos cognitivos centrais são não-modulares, isso é péssima notícia para as ciências cognitivas (FODOR, 1983, p. 128).

À luz de Fodor, então, considerações que militam contra a modularidade de nível superior também militam contra a possibilidade de uma ciência robusta da alta cognição, resultado infeliz, considerando os interesses da maior parte dos cientistas cognitivos e filósofos da mente.

Consequências sombrias à parte, é difícil oferecer resistência ao argumento de Fodor contra a modularidade de nível superior. Os principais pontos críticos são os seguintes: primeiro, a correlação negativa entre globalidade e encapsulamento; segundo, a correlação positiva entre encapsulamento e modularidade. Postos conjuntamente, temos uma correlação negativa entre globalidade e modularidade: quanto mais global o processo, menos modular é o sistema que o executa. Assim, parece haver apenas três caminhos para barrar a conclusão do argumento:

- 1 Negar que processos centrais sejam globais.
- 2 Negar que globalidade e encapsulamento sejam negativamente correlacionados.
- 3 Negar que encapsulamento e modularidade sejam positivamente correlacionados.

Dessas três opções, a segunda parece a menos atraente, na medida em que parece uma verdade conceitual que globalidade e encapsulamento andam em sentidos opostos. A primeira opção tem um apelo ligeiramente maior, mas apenas ligeiramente. A ideia de que processos centrais sejam relativamente globais, ainda que não tão globais quanto processos de confirmação científica sugerem, é difícil de negar. E isso é tudo que o argumento requer.

Isso nos deixa com a terceira opção: negar que a modularidade requer encapsulamento. É esta, com efeito, a estratégia usada por Carruthers (2006). Mais precisamente, Carruthers faz uma distinção entre dois tipos de encapsulamento: de ‘escopo estreito’ e de ‘escopo abrangente’. Um sistema é estreitamente encapsulado relativamente ao seu escopo se é incapaz de considerar **qualquer** informação externa ao próprio sistema no curso de sua operação. Isso corresponde ao encapsulamento tal como Fodor usa o termo. Por outro lado, um sistema que é abrangentemente encapsulado relativamente ao seu escopo pode operar sobre informações exógenas no decurso de suas operações, o sistema apenas não pode operar sobre **todas** essas informações. (Compare: “Nenhuma informação exógena é acessível” vs. “Algumas informações exógenas não são acessíveis”). Trata-se de encapsulamento num sentido mais fraco do termo do que o de Fodor. De fato, o uso de Carruthers, do termo ‘encapsulamento’, nesse contexto, é algo enganador, na medida em que sistemas abrangentemente encapsulados contam como não-encapsulados no sentido de Fodor (PRINZ, 2006).

Abandonar o requisito do encapsulamento (estrito) para módulos levanta uma série de questões, entre as não menos importantes das quais sendo que isso reduz o poder da hipótese da modularidade de explicar dissociações funcionais a nível de sistema (STOKES; BERGERON, 2015). Dito isto, se a modularidade requer apenas um encapsulamento abrangente, então o argumento de Fodor contra a modularidade central deixa de funcionar. Porém, dada a importância do encapsulamento estreito para a modularidade fodoriana, tudo que isso mostra é que sistemas centrais podem ser modulares de uma maneira não-fodoriana. O argumento original de que sistemas centrais não são modulares *à la* Fodor, e, com este, a motivação para a vertente negativa da hipótese da modularidade modesta, permanece de pé.

### 3. Modularidade pós-Fodor

De acordo com a hipótese da modularidade massiva, a mente é inteiramente modular, incluindo as partes responsáveis por funções cognitivas de alto nível, como a fixação de crenças, a solução de problemas, o planejamento e similares. Originalmente articulada e defendida por proponentes da psicologia evolutiva (SPERBER, 1994, 2002; COSMIDES; TOOBY, 1992; PINKER, 1997; BARRETT, 2005; BARRETT; KURZBAN, 2006), a hipótese recebeu a defesa mais abrangente e sofisticada pelas mãos de Carruthers (2006). Contudo, antes de proceder aos

detalhes dessa defesa, precisamos brevemente considerar que conceito de modularidade está em jogo.

O principal a se notar aqui é que a noção operante de modularidade difere significativamente daquela tradicional fodoriana. Carruthers é explícito neste ponto:

[Se] uma tese da modularidade mental massiva é para ser remotamente plausível, então por 'módulo' não podemos significar 'módulo *a la* Fodor'. Em particular, as propriedades de ter transdutores próprios, respostas rasas, processamentos rápidos, inatismo significativo ou canalização inata e encapsulamento terão muito provavelmente de ser descartadas. Isso nos deixa com a ideia de que módulos podem ser sistemas de processamento isoláveis e funcionalmente específicos, todos ou quase todos de domínio (no sentido do conteúdo), com operações insubordinadas à vontade consciente, associados a estruturas neurais específicas (ainda que em alguns casos dispersas espacialmente) e cujas operações internas podem ser inacessíveis para o restante da cognição. (CARRUTHERS, 2006, p. 12)

Do conjunto inicial de nove características associadas com os módulos de Fodor, portanto, os módulos *a la* Carruthers retêm, no máximo, apenas cinco: dissociabilidade, especificidade de domínio, automaticidade, localizabilidade neural e inacessibilidade central. Notadamente ausente da lista está o encapsulamento informacional, a característica mais central à modularidade na abordagem de Fodor. Ademais, Carruthers virá a dispensar a especificidade de domínio, a automaticidade e a localizabilidade forte (que não permite a sobreposição de módulos) de sua lista inicial de cinco características, tornando sua concepção de modularidade ainda mais esparsa (CARRUTHERS, 2006, P. 62). Outras propostas na literatura mostram-se semelhantemente permissivas em termos dos requisitos que um sistema precisa cumprir para contar como modular (COLTHEART, 1999; BARRETT; KURZBAN, 2006).

Um segundo ponto, relacionado ao primeiro, é que defensores da modularidade massiva têm tido como maior preocupação a defesa da modularidade da cognição central, tomando como certo que a mente é modular a nível dos sistemas de entrada. Assim, a hipótese em questão para teóricos como Carruthers pode ser melhor



entendida como a conjunção de duas alegações: em primeiro lugar, que sistemas de entrada são modulares de maneira a requerer encapsulamento de escopo estrito; em segundo lugar, que sistemas centrais são modulares, mas apenas de maneira a não requerer essa característica. Ao defender a modularidade massiva, Carruthers concentra-se na segunda dessas alegações, e assim nós o faremos.

### 3.1 O argumento em prol da modularidade massiva

A peça central de Carruthers (2006) consiste em três argumentos em prol da modularidade massiva: o Argumento do *Design*, o Argumento dos Animais e o Argumento da Tratabilidade Computacional. Consideremos brevemente cada um deles.

O Argumento do Design é como segue:

- 1 Sistemas biológicos são sistemas projetados, incrementalmente construídos.
- 2 Quando complexos, tais sistemas têm de ser organizados de maneira massivamente modular, isto é, como uma reunião hierárquica de componentes separadamente modificáveis e funcionalmente autônomos.
- 3 A mente humana é um sistema biológico, e é complexa.
- 4 Logo, a mente humana é (provavelmente) massivamente modular em sua organização. (CARRUTHERS, 2006, p. 25)

O ponto crucial desse argumento é a ideia de que sistemas biológicos complexos não podem evoluir se não estiverem organizados de maneira modular, sendo que a organização modular implica que cada componente do sistema (isto é, cada módulo) pode ser selecionado para mudança independentemente dos outros. Noutras palavras, a possibilidade de um sistema como um todo evoluir requer a possibilidade da evolução independente das suas partes. O problema com essa assunção se desdobra em duas partes (WOODWARD; COWIE, 2004). Primeiro, nem todos os traços biológicos são independentemente modificáveis. Ter dois pulmões, por exemplo, é um traço que não pode ser mudado sem que sejam mudados outros traços do organismo, pois os mecanismos genéticos e de desenvolvimento subjacentes ao número de pulmões dependem causalmente dos

mecanismos genéticos e desenvolvimentais subjacentes à simetria bilateral. Segundo, parece haver restrições desenvolvimentais sobre a neurogênese, as quais impedem que o tamanho de uma área cerebral se altere independentemente de alterações no tamanho de outras áreas do cérebro. Isso, por sua vez, indica que a seleção natural não pode modificar traços cognitivos em isolamento uns dos outros, dado que a evolução dos circuitos neurais de um traço cognitivo provavelmente resultará em alterações nos circuitos neurais de outros traços.

Um outro problema do Argumento do *Design* concerne à lacuna entre a sua conclusão (a alegação de que a mente é massivamente modular **quanto à organização**) e a hipótese em questão (a alegação de que a mente é massivamente modular *simpliciter*). O problema é este. De acordo com Carruthers, a modularidade de um sistema implica a posse de apenas duas propriedades: dissociabilidade funcional e inacessibilidade do processamento ao monitoramento externo. Suponha que um sistema é massivamente modular quanto à sua organização. Segue-se da definição de organização modular que os componentes do sistema são funcionalmente autônomos e separadamente modificáveis. Conquanto a autonomia funcional garanta a dissociabilidade, não é clara a razão pela qual a separabilidade das modificações garanta a inacessibilidade ao monitoramento externo. De acordo com Carruthers, a razão é que “se as operações internas de um sistema (por exemplo, os detalhes do algoritmo sendo executado) estivessem disponíveis alhures, então elas não poderiam ser alteradas sem que, correspondentemente, ocorressem alterações no sistema para o qual elas são acessíveis” (CARRUTHERS, 2006, P. 61). Mas essa é uma assunção questionável. Pelo contrário, parece plausível que as operações internas de um sistema poderiam ser acessíveis a um segundo sistema em virtude de um mecanismo de monitoramento que funciona do mesmo jeito à revelia dos detalhes do processamento monitorado. No mínimo, a alegação de que a modificabilidade em separado implica inacessibilidade ao monitoramento externo pede por mais justificação do que Carruthers oferece.

Em suma, o Argumento do *Design* é suscetível a várias objeções. Felizmente, há um argumento ligeiramente mais forte que lhe segue de perto, devido a Cosmides e Tooby (1992). É o seguinte:

- 1 A mente humana é um produto da seleção natural.
- 2 A fim de sobreviver e se reproduzir, nossos ancestrais humanos tiveram de resolver um sem-número de problemas adaptativos recorrentes (encontrar comida,

abrigo, parceiros, etc.).

- 3 Porque problemas adaptativos são resolvidos mais rápida, eficiente e confiavelmente por sistemas modulares do que por sistemas não-modulares, a seleção natural teria favorecido a evolução de uma arquitetura massivamente modular.
- 4 Logo, a mente humana é (provavelmente) massivamente modular.

A força desse argumento depende especialmente da força da terceira premissa. Nem todas as pessoas são convencidas, para dizer o mínimo (FODOR, 2000; SAMUELS, 2000; WOODWARD; COWIE, 2004). Primeiro, a premissa exemplifica o raciocínio adaptacionista, e o adaptacionismo na filosofia da biologia tem mais que a sua cota de críticas. Segundo, é duvidoso se a resolução de problemas adaptativos é em geral mais fácil de realizar com uma grande coleção de dispositivos especializados de resolução de problemas do que com uma coleção menor de dispositivos de resolução de problemas genéricos com acesso a uma biblioteca de programas especializados (SAMUELS, 2000). Sendo assim, na medida em que a hipótese da modularidade massiva postula uma arquitetura do primeiro tipo, como sugere a metáfora do ‘canivete suíço’ dos psicólogos evolucionistas (COSMIDES; TOOBY, 1992), a premissa parece instável.

Um argumento relacionado é o Argumento dos Animais. Diferente do que ocorre com o Argumento do Design, esse argumento nunca é explicitamente formulado em Carruthers (2006). Porém, eis uma de suas reconstruções plausíveis, devida a Wilson (2008):

- 1 Mentes animais são massivamente modulares.
- 2 Mentes humanas são extensões incrementais de mentes animais.
- 3 Logo, a mente humana é (provavelmente) massivamente modular.

Infelizmente para aqueles que defendem a modularidade massiva, este argumento, assim como o argumento do design, é vulnerável a várias objeções (WILSON, 2008). Mencionaremos duas delas. Primeiro, não é fácil motivar a alegação de que mentes animais são massivamente modulares no sentido operante. Ainda que Carruthers (2006) realize um esforço heroico para fazê-lo, as evidências que

ele cita, por exemplo, em favor da especificidade de domínio dos mecanismos animais de aprendizado, *à la* Gallistel (1990), somam menos do que seria necessário. O problema é que a especificidade de domínio não basta para se ter a modularidade *à la* Carruthers; de fato, não é sequer uma das características centrais da modularidade na explicação de Carruthers. Portanto, o argumento tropeça no primeiro passo. Segundo, mesmo se mentes animais forem massivamente modulares, e mesmo se extensões incrementais singulares da mente animal preservem este traço, é bem possível que uma série de tais extensões de mentes animais possam ter levado à sua perda. Em outras palavras, como pontua Wilson (2008), não se pode assumir que a conservação da modularidade massiva é transitiva. E, sem essa assunção, o argumento dos animais não se sustenta.

Finalmente, temos o argumento da tratabilidade computacional (CARRUTHERS, 2006, p. 44-59). Para os propósitos desse argumento, assumimos que um processo mental é computacionalmente tratável se ele puder ser especificado o nível algorítmico de tal modo que a execução do processo seja viável em certo tempo, com certo dispêndio energético e com outras restrições materiais que incidem sobre a cognição humana (SAMUELS, 2005). Assumimos também que um sistema é encapsulado se no decurso de suas operações o sistema carece de acesso a alguma informação exógena.

- 1 A mente é computacionalmente realizada.
- 2 Todos os processos mentais computacionais devem ser tratáveis.
- 3 Processamentos tratáveis são possíveis somente em sistemas encapsulados.
- 4 Logo, a mente deve consistir inteiramente de sistemas encapsulados.
- 5 Logo, a mente é (provavelmente) massivamente modular.

Existem dois problemas com este argumento, no entanto. O primeiro problema tem a ver com a terceira premissa, que afirma que a tratabilidade requer encapsulamento, isto é, a inacessibilidade a pelo menos alguma informação exógena ao processamento. O que a tratabilidade de fato requer é algo mais fraco, nomeadamente, que nem toda informação seja acessada pelo mecanismo no decurso de suas operações (SAMUELS, 2005). Em outras palavras, é possível que um sistema tenha acesso ilimitado a uma base de dados sem que efetivamente

acesse todos os seus conteúdos. Apesar de a computação tratável excluir a pesquisa exaustiva, por exemplo, mecanismos não-encapsulados não precisam se engajar em pesquisas exaustivas; assim, a tratabilidade não requer o encapsulamento. O segundo problema com o argumento é relativo ao último passo. Ainda que alguém possa supor que sistemas modulares devem ser encapsulados, o inverso não se segue. De fato, Carruthers (2006) não faz menção ao encapsulamento na sua caracterização da modularidade, então não fica claro como se deve passar de uma alegação acerca do encapsulamento pervasivo para uma alegação acerca da modularidade pervasiva.

Resumindo, então, argumentos gerais convincentes em prol da modularidade massiva são difíceis de encontrar. Isso não importa ainda em abandonar a possibilidade de modularidade na cognição superior, mas abre alas ao ceticismo, especialmente em razão da penúria de evidências empíricas em apoio direto à hipótese (ROBBINS, 2013). Por exemplo, já foi sugerido que a capacidade de pensar sobre trocas sociais é devida a um mecanismo inato, funcionalmente dissociável e de domínio específico (STONE *et al.*, 2002; SUGIYAMA *et al.*, 2002). No entanto, o que parece é que perdas de razoabilidade em trocas sociais não ocorrem em isolamento, mas são acompanhadas por outros comprometimentos cognitivo-sociais (PRINZ, 2006). O ceticismo sobre a modularidade em outras áreas da cognição central, tais como a da leitura mental de alto nível, também parece prevalecer (CURRIE; STERELNY, 2000). O tipo de comprometimento da leitura mental característico da síndrome de Asperger e de graus elevados de autismo, por exemplo, ocorrem concomitantemente a perdas nas capacidades de processamento sensorial e de função executiva (FRITH, 2003). Em geral, não há muito em termos de evidência neuropsicológica em apoio à ideia de modularidade de alto nível.

### 3.2 Dúvidas sobre a modularidade massiva

Assim como há argumentos gerais e abstratos em prol da modularidade massiva, há também argumentos gerais e abstratos contra ela. Um destes toma a forma daquilo a que Fodor (2000) chama o 'Problema do *Input*'. O problema é este. Suponha que a arquitetura da mente é completamente modular, e que a mente consiste inteiramente de mecanismos de domínio específico. Neste caso, as respostas de cada sistema (de entrada) de baixo nível terão de ser roteadas para o sistema superior apropriadamente especializado (central) para processamento ulterior. Mas

esse roteamento só pode ser feito por um mecanismo não-modular, de domínio generalizado, contradizendo a suposição inicial. Em resposta a esse problema, Barrett (2005) argumenta que o processamento em uma arquitetura massivamente modular não requer um dispositivo de roteamento de domínio geral do tipo imaginado por Fodor. Barrett sugere que uma alternativa envolveria algo a que ele chama 'computação enzimática'. Nesse modelo, sistemas de baixo nível reúnem suas respostas em uma área de trabalho acessível a sistemas centrais, onde cada um destes é seletivamente ativado por respostas que satisfazem as especificidades de seu domínio, basicamente da mesma maneira que enzimas se ligam seletivamente aos substratos que lhes são especificamente adequados. Como enzimas, dispositivos computacionais especializados a nível central da arquitetura aceitam somente uma gama restrita de entradas (análogas a substratos bioquímicos), performam operações especializadas (análogas às reações bioquímicas) sobre essas entradas e, por fim, produzem respostas (análogas a produtos bioquímicos) em um formato utilizável por outros dispositivos computacionais. Isso elimina a necessidade de haver um mecanismo de domínio geral (portanto, não-modular) para fazer a mediação entre sistemas de baixo e alto níveis.

Um segundo desafio à modularidade massiva advém do 'problema da integração de domínios' (CARRUTHERS, 2006). O problema aqui é que o raciocínio, o planejamento, a tomada de decisões e outros tipos de cognição superior envolvem, rotineiramente, a produção de representações conceitualmente estruturadas cujos conteúdos atravessam domínios. Isso significa que deve haver algum mecanismo para integrar representações de múltiplos domínios. Porém, esse mecanismo teria de ser de domínio geral, e não específico, e portanto não seria modular. Assim como o Problema do *Input*, no entanto, o Problema da Integração de Domínios não é intransponível. Uma possível solução é que o sistema da linguagem tem a capacidade de desempenhar o papel de integrador de conteúdos em virtude de sua capacidade de transformar representações conceituais que tenham sido linguisticamente codificadas (HERMER; SPELKE, 1996; CARRUTHERS, 2002, 2006). Sob esse ponto de vista, a linguagem é o veículo para o pensamento de domínio geral.

Objecções empíricas à modularidade massiva assumem diversas formas. Para começar, existem evidências neurobiológicas de plasticidade de desenvolvimento, um fenômeno contrário à ideia de que a estrutura cerebral é especificada de maneira inata (BULLER, 2005; BULLER; HARDCASTLE, 2000). No entanto, nem todos os proponentes da modularidade massiva insistem que modos são especificados de maneira inata (CARRUTHERS, 2006; KURZBAN; TOOBY; COSMIDES, 2001).

Ademais, não está claro em que medida o registro neurobiológico conflita com o inatismo, dada a evidência de que genes específicos estão ligados ao desenvolvimento normal de estruturas corticais tanto em humanos quanto em animais (MACHERY; BARRETT, 2008; RAMUS, 2006).

Uma outra fonte de evidência contra a modularidade massiva vem de pesquisas sobre diferenças individuais na cognição superior (RABAGLIA; MARCUS; LANE, 2011). Tais diferenças tendem a ser fortemente correlacionadas de modo positivo entre vários domínios, um fenômeno conhecido como o 'arcabouço positivo', sugerindo que habilidades cognitivas de alto nível subordinam mecanismos de domínio geral, ao invés de um aparato de módulos especializados. Há, no entanto, uma explicação alternativa ao arcabouço positivo. Porque aos módulos pós-Fodor é permitido o compartilhamento de partes (CARRUTHERS, 2006), as correlações observadas podem se dever a diferenças individuais no funcionamento de componentes ao longo de múltiplos mecanismos de domínios específicos.

#### **4. Modularidade e a filosofia**

O interesse pela modularidade não está confinado às ciências cognitivas e à filosofia da mente, mas se estende por diversos campos cooperantes. Em epistemologia, a modularidade já foi evocada para defender a legitimidade de uma noção de observação neutra com respeito à teoria, e portanto a possibilidade de algum grau de consenso entre cientistas com comprometimentos teóricos divergentes (FODOR, 1984). O debate corrente sobre esse assunto (CHURCHLAND, 1988; FODOR, 1988; MCCAULEY; HENRICH, 2006) mantém uma significância duradoura para a filosofia da ciência em geral, particularmente em relação a controvérsias em torno do status do realismo científico. De modo relacionado, evidências em prol da penetrabilidade cognitiva da percepção trouxeram à tona questionamentos sobre a justificação de crenças perceptuais (SIEGEL, 2012; STOKES, 2012). Na ética, evidências desse tipo foram usadas para lançar dúvidas sobre o intuicionismo ético como abordagem à epistemologia moral (COWAN, 2014). Na filosofia da linguagem, a modularidade tem figurado em teorias acerca da comunicação linguística, por exemplo, na sugestão de teóricos da relevância de que a interpretação da fala, com seus aspectos pragmáticos e tudo mais, seria um processo modular (SPERBER; WILSON, 2002). Também se serviu da modularidade para demarcar a fronteira entre a semântica e a pragmática, e para defender uma versão notavelmente austera

de minimalismo semântico (BORG, 2004). Ainda que o sucesso desses empregos da teoria da modularidade esteja sujeito a disputas (por exemplo, ver ROBBINS, 2007, para dúvidas acerca da modularidade da semântica), sua existência testemunha a favor da relevância do conceito de modularidade para a investigação filosófica em uma variedade de domínios.

### Referência bibliográfica

- ANDERSON, M. Neural reuse: A fundamental organizational principle of the brain. **Behavioral and Brain Sciences**, v. 33, p. 245-313, 2010.
- ANDERSON, M. **After phrenology: Neural reuse and the interactive brain**. Cambridge, MA: MIT Press, 2014.
- ANTONY, L. Rabbit-pots and supernovas: On the relevance of psychological data to linguistic theory. *In*: BARBER, A. (ed.). **Epistemology of Language**. Oxford: Oxford University Press, 2003. p. 47-68.
- ARBIB, M. Modularity and interaction of brain regions underlying visuomotor coordination. *In*: GARFIELD, J. (ed.). **Modularity in Knowledge Representation and Natural-Language Understanding**. Cambridge, MA: MIT Press, 1987. p. 333-363.
- ARIEW, A. Innateness is canalization: In defense of a developmental account of innateness. *In*: HARDCASTLE, V. (ed.). **Where Biology Meets Psychology**. Cambridge, MA: MIT Press, 1999. p. 117-138.
- BALCETIS, E. Wishful seeing: More desired objects are seen as closer. **Psychological Science**, v. 21, p. 147-152, 2010.
- BALCETIS, E.; DUNNING, D. See what you want to see: Motivational influences on visual perception. **Journal of Personality and Social Psychology**, v. 91, p. 612-625, 2006.
- BARGH, J.; CHARTRAND, T. The unbearable automaticity of being. **American Psychologist**, v. 54, p. 462-479, 1999.
- BARRETT, H. Enzymatic computation and cognitive modularity. **Mind & Language**, v. 20, p. 259-287, 2005.
- BARRETT, H.; KURZBAN, R. Modularity in cognition: Framing the debate. **Psychological Review**, v. 113, p. 628-647, 2006a.
- BARRETT, H.; KURZBAN, R. Modularity in cognition: Framing the debate. **Psychological Review**, v. 113, p. 628-647, 2006b.



- BORG, E. **Minimal Semantics**. Oxford: Oxford University Press, 2004.
- BRUNER, J.; GOODMAN, C. Value and need as organizing factors in perception. **Journal of Abnormal and Social Psychology**, v. 42, p. 33-44, 1947.
- BULLER, D. **Adapting Minds**. Cambridge, MA: MIT Press, 2005.
- BULLER, D.; HARDCASTLE, V. Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. **Brain and Mind**, v. 1, p. 302-325, 2000.
- CARRUTHERS, P. The cognitive functions of language. **Behavioral and Brain Sciences**, v. 25, p. 657-725, 2002.
- CARRUTHERS, P. **The Architecture of the Mind**. Oxford: Oxford University Press, 2006.
- CHURCHLAND, P. Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. **Philosophy of Science**, v. 55, p. 167-187, 1988.
- COLTHEART, M. Modularity and cognition. **Trends in Cognitive Sciences**, v. 3, p. 115-120, 1999.
- COSMIDES, L.; TOOBY, J. Cognitive adaptations for social exchange. In: BARKOW, J.; COSMIDES, L.; TOOBY, J. (ed.). **The Adapted Mind**. Oxford: Oxford University Press, 1992. p. 163-228.
- COWAN, R. Cognitive penetrability and ethical perception. **Review of Philosophy and Psychology**, v. 6, p. 665-682, 2014.
- COWIE, F. **What's Within? Nativism Reconsidered**. Oxford: Oxford University Press, 1999.
- CURRIE, G.; STERELNY, K. How to think about the modularity of mind-reading. **Philosophical Quarterly**, v. 50, p. 145-160, 2000.
- FIRESTONE, C.; SCHOLL, B. Cognition does not affect perception: Evaluating the evidence for "top-down" effects. **Behavioral and Brain Sciences**, v. 39, 2016.
- FODOR, J. **The Modularity of Mind**. Cambridge, MA: MIT Press, 1983.
- FODOR, J. Observation reconsidered. **Philosophy of Science**, v. 51, p. 23-43, 1984.
- FODOR, J. A reply to Churchland's "Perceptual plasticity and theoretical neutrality". **Philosophy of Science**, v. 55, p. 188-198, 1988.
- FODOR, J. **The Mind Doesn't Work That Way**. Cambridge, MA: MIT Press, 2000.
- FRITH, U. **Autism: Explaining the enigma**. Malden, MA: Wiley-Blackwell, 2003.
- GAUTHIER, I. *et al.* Expertise for cars and birds recruits brain areas involved in face recognition. **Nature Neuroscience**, v. 3, p. 191-197, 2000.
- HANSEN, T. *et al.* Memory modulates color appearance. **Nature Neuroscience**, v. 9, p. 1367-1368, 2006.

- HERMER, L.; SPELKE, E. Modularity and development: The case of spatial reorientation. **Cognition**, v. 61, p. 195-232, 1996.
- KURZBAN, R.; TOOBY, J.; COSMIDES, L. Can race be erased? Coalitional computation and social categorization. **Proceedings of the National Academy of Sciences**, v. 98, p. 15387-15392, 2001.
- LEVIN, D.; BANAJI, M. Distortions in the perceived lightness of faces: The role of race categories. **Journal of Experimental Psychology: General**, v. 135, p. 501-512, 2006.
- MACHERY, E. Cognitive penetrability: A no-progress report. *In*: ZEIMBEKIS, J.; RAFTOPOULOS, A. (ed.). **The Cognitive Penetrability of Perception: New Philosophical Perspectives**. Oxford: Oxford University Press, 2015.
- MACHERY, E.; BARRETT, H. Debunking Adapting Minds. **Philosophy of Science**, v. 73, p. 232-246, 2006.
- MARSLEN-WILSON, W.; TYLER, L. Against modularity. *In*: GARFIELD, J. (ed.). **Modularity in Knowledge Representation and Natural-Language Understanding**. Cambridge, MA: MIT Press, 1987.
- MCCAULEY, R.; HENRICH, J. Susceptibility to the Müller-Lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. **Philosophical Psychology**, v. 19, p. 79-101, 2006.
- MCGURK, H.; MACDONALD, J. Hearing lips and seeing voices. **Nature**, v. 391, p. 756, 1976.
- PINKER, S. **How the Mind Works**. New York: W. W. Norton & Company, 1997.
- PRINZ, J. Is the mind really modular? *In*: STANTON, R. (ed.). **Contemporary Debates in Cognitive Science**. Oxford: Blackwell, 2006. p. 22-36.
- PYLYSHYN, Z. **Computation and Cognition**. Cambridge, MA: MIT Press, 1984.
- PYLYSHYN, Z. Is vision continuous with cognition? The case for cognitive penetrability of vision. **Behavioral and Brain Sciences**, v. 22, p. 341-423, 1999.
- RABAGLIA, C.; MARCUS, G.; LANE, S. What can individual differences tell us about the specialization of function? **Cognitive Neuropsychology**, v. 28, p. 288-303, 2011.
- RAMUS, F. Genes, brain, and cognition: A roadmap for the cognitive scientist. **Cognition**, v. 101, p. 247-269, 2006.
- ROBBINS, P. Minimalism and modularity. *In*: PREYER, G.; PETER, G. (ed.). **Context-Sensitivity and Semantic Minimalism**. Oxford: Oxford University Press, 2007. p. 303-319.

- ROBBINS, P. Modularity and mental architecture. **WIREs Cognitive Science**, v. 4, p. 641-649, 2013.
- ROSCH, E. *et al.* Basic objects in natural categories. **Cognitive Psychology**, v. 8, p. 382-439, 1976.
- SAMUELS, R. Massively modular minds: Evolutionary psychology and cognitive architecture. *In*: CARRUTHERS, P.; CHAMBERLAIN, A. (ed.). **Evolution and the Human Mind**. Cambridge: Cambridge University Press, 2000. p. 13-46.
- SAMUELS, R. The complexity of cognition: Tractability arguments for massive modularity. *In*: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (ed.). **The Innate Mind: Structure and Contents**. Oxford: Oxford University Press, 2005. p. 107-121.
- SCHOLL, B.; LESLIE, A. Modularity, development and "theory of mind". **Mind & Language**, v. 14, p. 131-153, 1999.
- SEGAL, G. The modularity of theory of mind. *In*: CARRUTHERS, P.; SMITH, P. (ed.). **Theories of Theories of Mind**. Cambridge: Cambridge University Press, 1996. p. 141-157.
- SEGALL, M.; CAMPBELL, D.; HERSKOVITS, M. **The Influence of Culture on Visual Perception**. New York: Bobbs-Merrill, 1966.
- SHAMS, L.; KAMITANI, Y.; SHIMOJO, S. Illusions: What you see is what you hear. **Nature**, v. 408, p. 788, 2000.
- SIEGEL, S. Cognitive penetrability and perceptual justification. **Nous**, v. 46, p. 201-222, 2011.
- SPELKE, E. Initial knowledge: Six suggestions. **Cognition**, v. 50, p. 435-445, 1994.
- SPERBER, D. The modularity of thought and the epidemiology of representations. *In*: HIRSCHFELD, L.; GELMAN, S. (ed.). **Mapping the Mind**. Cambridge: Cambridge University Press, 1994. p. 39-67.
- SPERBER, D. In defense of massive modularity. *In*: DUPOUX, I. (ed.). **Language, Brain, and Cognitive Development**. Cambridge, MA: MIT Press, 2002. p. 47-57.
- SPERBER, D.; WILSON, D. Pragmatics, modularity and mind-reading. **Mind & Language**, v. 17, p. 3-23, 2002.
- STOKES, D. Perceiving and desiring: A new look at the cognitive penetrability of experience. **Philosophical Studies**, v. 158, p. 479-492, 2012.
- STOKES, D. Cognitive penetrability of perception. **Philosophy Compass**, v. 8, p. 646-663, 2013.

- STOKES, D.; BERGERON, V. Modular architectures and informational encapsulation: A dilemma. **European Journal for the Philosophy of Science**, v. 5, p. 315-338, 2015.
- STONE, V. *et al.* Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. **Proceedings of the National Academy of Sciences**, v. 99, p. 11531-11536, 2002.
- STROMSWOLD, K. Cognitive and neural aspects of language acquisition. *In*: LEPORE, E.; PYLYSHYN, Z. (ed.). **What Is Cognitive Science?** Oxford: Blackwell, 1999. p. 356-400.
- SUGIYAMA, L.; TOOBY, J.; COSMIDES, L. Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. **Proceedings of the National Academy of Sciences**, v. 99, p. 11537-11542, 2002.
- TETTAMANTI, M.; WENIGER, D. Broca's area: A supramodal hierarchical processor? **Cortex**, v. 42, p. 491-494, 2006.
- WARREN, R. Perceptual restoration of missing speech sounds. **Science**, v. 167, p. 392-393, 1970.
- WARREN, R.; WARREN, R. Auditory illusions and confusions. **Scientific American**, v. 223, p. 30-36, 1970.
- WILSON, R. The drink you're having when you're not having a drink. **Mind & Language**, v. 23, p. 273-283, 2008.
- WITT, J. *et al.* Putting to a bigger hole: Golf performance relates to perceived size. **Psychonomic Bulletin and Review**, v. 15, p. 581-585, 2008.
- WITT, J.; PROFFITT, D.; EPSTEIN, W. Perceiving distances: A role of effort and intent. **Perception**, v. 33, p. 577-590, 2004.
- WOODS, A.; PHILBECK, J.; DANOFF, J. The various perceptions of distance: An alternative view of how effort affects distance judgments. **Journal of Experimental Psychology: Human Perception and Performance**, v. 35, p. 1104-1117, 2009.
- WOODWARD, J.; COWIE, F. The mind is not (just) a system of modules shaped (just) by natural selection. *In*: HITCHCOCK, C. (ed.). **Contemporary Debates in Philosophy of Science**. Malden, MA: Blackwell, 2004. p. 312-334.

## (V) Cognição Corporificada\*

Autor: Robert A. Wilson e Lucia Foglia  
Tradução Rodrigo Ulhôa (PPGFIL-UFRGS)  
Revisão: César Fernando Meurer (UFABC)

A cognição é corporificada quando depende profundamente de recursos do corpo físico de um agente, isto é, quando aspectos do corpo do agente, além do cérebro, desempenham uma função causal significativa ou fisicamente constitutiva no processo cognitivo.

Em geral, posições dominantes em filosofia da mente e em ciência cognitiva consideraram o corpo como periférico para compreender a natureza da mente e da cognição. Os proponentes da ciência cognitiva corporificada observam que isso é um grave engano. Às vezes a natureza da dependência da cognição em relação ao corpo é bastante inesperada, e sugere novas formas de conceitualizar e explorar

---

\*WILSON, R. A.; FOGLIA, L. Embodied Cognition. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2017. Disponível em: <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/>. Acesso em: 30 out. 2021.

The following is the translation of the entry on Embodied Cognition by Robert A. Wilson and Lucia Foglia, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/embodied-cognition/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.

as mecânicas do processamento cognitivo.

A ciência cognitiva corporificada abarca uma família tenuemente conectada de programas de pesquisa nas ciências cognitivas que frequentemente compartilham um compromisso de criticar e até mesmo de substituir abordagens tradicionais da cognição e do processamento cognitivo. As pesquisas empíricas sobre a cognição corporificada explodiram nos últimos 10 anos. Como atesta a bibliografia desse artigo, os vários textos que serão discutidos representam uma promissora alternativa à investigação de fenômenos cognitivos.

O trabalho relativamente recente sobre a corporificação da cognição oferece muito alimento para o pensamento dos filósofos da mente que são empiricamente informados. Em parte, isso se deve à rica gama de fenômenos que a ciência cognitiva corporificada estuda. Em parte, é também porque esses fenômenos são frequentemente tomados como desafiando as posições dominantes a respeito da mente, tal como as teorias computacionais e representacionais, que estão no coração da ciência cognitiva tradicional. Tais fenômenos foram em alguns casos usados para minar posições padrão em filosofia da mente, como a ideia de que a mente é idêntica ao, ou mesmo realizada no cérebro.

## 1. Ciência Cognitiva Tradicional vs. Ciência Cognitiva Corporificada

Considere quatro exemplos evocativos de fenômenos que motivaram a ciência cognitiva corporificada:

- 1 Nós tipicamente gesticulamos quando falamos com alguém, e gesticular facilita não apenas a comunicação, mas o próprio processamento da linguagem (MCNEILL, 1992).
- 2 A visão é em geral orientada para a ação, e o movimento corpóreo e a reação que ela gera são mais fortemente integrados em ao menos algum processo visual do que os modelos tradicionais da visão anteciparam (O'REGAN; NOË, 2001).
- 3 Existem neurônios, *neurônios espelho*, que disparam não somente quando realizamos uma ação, mas também quando observamos outros realizando as mesmas ações (RIZZOLATTI; CRAIGHERO, 2004).

- 4 Somos frequentemente capazes de realizar tarefas cognitivas de modo mais efetivo, como lembrar, usando nossos corpos e até mesmo partes de nosso ambiente circundante para armazenar e simplificar a natureza do processamento cognitivo (DONALD, 1991).

Embora fenômenos tais como (1)-(4) motivem a ciência cognitiva corporificada, apelar a esses fenômenos para chegar a conclusões mais substantivas do que as que foram esboçadas – por exemplo, que a ciência cognitiva tradicional é profundamente defeituosa, ou que concepções dominantes na filosofia da mente, tal como o funcionalismo, são equivocadas – exigem mais argumentos filosóficos. É porque a argumentação exigida tipicamente apela a outros conceitos centrais no trabalho em filosofia da mente empiricamente informada e na própria ciência cognitiva, tal como a modularidade e o nativismo, que o debate sobre a cognição corporificada se tornou um tópico quente na ciência cognitiva dos últimos anos (ADAMS, 2010; AIZAWA, 2007; CHEMERO, 2009; SHAPIRO 2011).

A ciência cognitiva tradicional certamente conceitualizou o processamento cognitivo central, o que chamaremos de cognição *no sentido estrito* [*in the narrow sense*], em abstração dos mecanismos corpóreos do processamento sensorial e do controle motor. Programas de pesquisa dentro da inteligência artificial exemplificam essa concepção da cognição no sentido estrito, e eles têm sido alvos mais claros da ciência cognitiva corporificada. Mais categoricamente, a ciência cognitiva corporificada visa compreender toda a gama de capacidades perceptuais, cognitivas e motoras que possuímos – a cognição **no sentido amplo** [*in the broad sense*] – como capacidades que são dependentes de recursos do corpo físico. Nesse artigo, consideramos a cognição construída tanto no sentido estrito quanto no sentido amplo.

Finalmente, como modo de introduzir a ciência cognitiva corporificada, observamos seu relacionamento com a *cognição situada* (SMITH, 1999; ROBBINS; AYDEDE, 2009). Como um paradigma dentro da cognição situada, a ciência cognitiva corporificada pode ser diferenciada tanto do estudo da cognição **situada** quanto da tese da cognição **estendida**.

A ciência cognitiva corporificada apela à ideia de que a cognição depende profundamente de aspectos do corpo, para além do cérebro do agente. Sem o envolvimento do corpo tanto em sentir quanto em agir, pensamentos seriam vazios e estados mentais não exibiriam as características e propriedades que eles exibem. O trabalho na cognição situada, por contraste, baseia-se na visão de que a cognição

depende profundamente do ambiente natural e social. Ao se concentrar nas estratégias que organismos usam para descarregar o processamento cognitivo sobre o ambiente, esse trabalho coloca ênfase particular nas maneiras pelas quais a atividade cognitiva é distribuída pelo agente e seu meio físico, social e cultural. (SUCHMAN, 1987; HUTCHINS, 1995). A tese da cognição estendida é a afirmação de que os próprios sistemas cognitivos se estendem para além dos limites do organismo individual. Dessa perspectiva, aspectos do ambiente físico, social e cultural de um agente podem fazer mais do que distribuir o processamento cognitivo: eles podem parcialmente constituir o sistema cognitivo desse agente. (CLARK & CHALMERS, 1998; WILSON, 2004; CLARK, 2008; MENARY, 2010).

Seguimos autores recentes (CLARK, 2008; RUPERT, 2009b; SHAPIRO, 2010, 2011) ao sustentar que, embora a ciência cognitiva corporificada possa ser nitidamente distinguida, em princípio, de outras formas de cognição situada, e que há momentos em que isso é útil (ou mesmo crucial), as questões filosóficas mais amplas em jogo também são às vezes discutidas de forma reveladora ao considerar essas concepções juntas. Assim, embora esse artigo se concentre em modos específicos pelos quais a cognição depende do corpo físico, ele também discute a cognição situada de modo mais geral, quando oportuno.

## 2. Algumas Bases Históricas da Ciência Cognitiva Corporificada

Uma consideração (Seções 2.1-2.3) sobre três publicações marcantes oferece uma base histórica para compreender os trabalhos iniciais em cognição corporificada no sentido estrito: *Metaphors We Live By* (1980), de George Lakoff e Mark Johnson, a perspectiva enativista sobre a cognição desenvolvida por Francisco Varela, Evan Thompson e Eleanor Rosch em *The Embodied Mind* (1991) e o trabalho sobre robótica e ação computacionalmente inteligente resumido e analisado por Andy Clark em *Being There: Putting Mind, World, and Body Back Together* (1997). Em seguida, retornamos brevemente às obras influentes da cognição corporificada no sentido amplo (Seções 2.4-2.5) e à tradição fenomenológica dentro da filosofia continental que inspirou a mais recente ciência cognitiva corporificada (Seção 2.6).



## 2.1 Metáfora e Cognição

A linguagem figurativa claramente desempenha um papel na cognição, e filósofos, linguistas e psicólogos têm contribuído para a compreensão desse papel na ciência cognitiva (BLACK, 1962; ORTONY, 1979). Logo no início de *Metaphors We Live By* (1980), George Lakoff e Mark Johnson argumentaram que tal linguagem, e a metáfora em particular, não era apenas um fenômeno que a ser estudado no domínio da cognição, mas que ativamente estrutura muito da cognição tradicionalmente considerada isolada da metáfora. Por exemplo, muitos dos processos cognitivos centrais, tais como aqueles relativos ao espaço e ao tempo, foram expressos e influenciados, dizem Lakoff e Johnson, por metáforas (assim, “metáforas pelas quais vivemos”). Se a experiência humana está intrinsecamente ligada a metáforas de grande escala, e tanto a experiência quanto a metáfora são moldadas pelos tipos de corpos que temos, que se interpõe entre o agente e o mundo, defendem Lakoff e Johnson, então a cognição é corporificada de uma maneira não antecipada dentro da ciência cognitiva tradicional.

Embora Lakoff (1987) e Johnson (1987) desenvolvam a ideia básica aqui de diferentes formas (*vide* LAKOFF; JOHNSON, 1999; JOHNSON, 2007), o tom geral da concepção que eles compartilham pode ser transmitido pela consideração de um exemplo bem conhecido que eles discutem: o do amor como um tipo de jornada. Frequentemente, diz-se que aqueles que estão num relacionamento romântico partem juntos, viajam pelo mesmo caminho, tomam o retorno errado, refazem seus passos, verificam seus rumos e fazem suas malas. Para Lakoff e Johnson, essa linguagem não-literal não é meramente uma expressão periférica útil por acrescentar sinos e assobios ao alvoroço da comunicação, mas reflete alguma coisa profunda sobre como o amor é conceitualizado. Principalmente, a metáfora organizadora central – o amor é uma jornada – envolve um mapear de um domínio (jornadas) a outro (amor), onde o domínio de origem é informado pela nossa fisicalidade corpórea e pela experiência corporificada que temos enquanto criaturas que se movem pelo mundo para realizar propósitos e objetivos.

Conceitos espaciais, tais como “à frente”, “atrás”, “acima” e “abaixo”, oferecem talvez os exemplos mais claros nos quais tal experiência corporificada existe. Esses conceitos são articulados em termos da posição e do movimento do nosso corpo no espaço. Criaturas como nós, que ficam de pé e se movem para a frente, pensam, por exemplo, sobre as coisas que estão “à frente” delas mesmas como localizadas na linha da visão ou em termos da direção em que estão se

movendo. Criaturas que fossem longas, planas e se movessem para trás, por contraste, poderiam ter um conceito muito diferente de “a frente de”, ou, talvez, nenhum conceito. Isso vale, também, para outros conceitos espaciais, como “acima”. Podemos sentir em primeira mão a natureza corporificada destes conceitos em situações em que nos aproximamos de tais criaturas, como quando tentamos usar esses conceitos para orientar nossa ação quando estamos deitados, nos movendo para trás, ou olhando no espelho retrovisor. Os proponentes da cognição corporificada alegam que a experiência de “acimidade” [*upness*] depende do tipo particular de corpo que temos e de como esse corpo interage com o seu entorno (LAKOFF; JOHNSON, 1999).

Esse exemplo também pode ser usado para ilustrar porque a cognição corporificada mostra ser uma posição controversa dentro da ciência cognitiva e da filosofia da mente. Pode-se argumentar que a dependência identificada acima, entre nossos conceitos espaciais e nossos corpos, é mundana e ordinária. Considere o mais conhecido dualista clássico mente-corpo, René Descartes. Em seu famoso gracejo na VI Meditação, de que ele (sua mente, para Descartes) não estava meramente alojado em seu corpo “como um piloto em um navio”, Descartes claramente reconhece que há algum sentido no qual a cognição depende do corpo e é integrada a ele. Os proponentes da cognição corporificada devem, no mínimo, mostrar que “à frente” e “acima” dependem do corpo em algum sentido que Descartes negaria.

De mais relevância para os debates contemporâneos, os proponentes da cognição corporificada devem mostrar que essa dependência não pode ser acomodada na ciência cognitiva tradicional e em seus compromissos de trabalho, por exemplo, com as teorias computacionais e representacionais da mente. Ao menos esse é o caso se a cognição corporificada for marcar um afastamento significativo da ciência cognitiva tradicional e das posições filosóficas associadas sobre a mente, colocando um desafio substancial a elas. E isto exige, pelo menos, não apenas identificar algum tipo de dependência entre a cognição e o corpo, mas especificar a natureza desta dependência.

## 2.2 Cognição Enativa

O Livro *The Embodied Mind* (VARELA; THOMPSON; ROSCH, 1991) foi uma tentativa de redirecionar as ciências cognitivas infundindo-as com a perspectiva fenomenológica desenvolvida na obra de Maurice Merleau-Ponty (1945). De modo

mais ambicioso, e menos bem-sucedido, ele também busca integrar a ciência cognitiva à filosofia Budista; o livro também inclui algumas discussões passageiras sobre psicanálise. Varela, Thompson e Rosch argumentaram que a divisão padrão entre aspectos pré-dados do mundo externo e representações simbólicas internas deveria ser abandonada, uma vez que ela é incapaz de acomodar o feedback das ações corporificadas à cognição a partir das ações de um agente cognitivo situado. As diferenças fundamentais entre as suas perspectivas e as da concepção clássica repousam nas respostas às questões sobre o que a cognição é, como ela opera e quando um sistema funciona adequadamente.

Concepções tradicionais basicamente afirmam que não há computações sem representações, e que a cognição funciona de modo bem-sucedido quando algum dispositivo pode suportar e manipular símbolos para solucionar o problema dado ao sistema. Varela, Thompson e Rosch introduziram o conceito de **enação** para apresentar e desenvolver um modelo que coloca forte ênfase na ideia de que o mundo experienciado é retratado e determinado por interações mútuas entre a fisiologia do organismo, seus circuitos sensório-motores e o ambiente. Suas ênfases sobre o acoplamento estrutural entre cérebro-corpo-mundo constitui o núcleo de seu programa de cognição corporificada, construído sobre a ideia da fenomenologia clássica de que agentes cognitivos realizam um mundo por meio das atividades de seus **corpos vivos situados**. Como a metáfora de “realizar um mundo” de experiências significativas implica, nessa visão o conhecimento emerge através do primário engajamento corporal do agente em um ambiente, ao invés de ser simplesmente determinado e dependente, seja de situações pré-existentes, seja de construções pessoais.

Uma consequência dessa visão é que apenas uma criatura com certas características, por exemplo, olhos, mãos, pernas e habilidades, pode possuir certos tipos de capacidades cognitivas. Isso porque a cognição é uma atividade sensório-motora dinâmica e o mundo que é dado e experienciado não é apenas condicionado pela atividade neural do sujeito, mas é essencialmente enativado na medida em que ele emerge pelas atividades corpóreas do organismo. Essa abordagem geral encoraja uma concepção de enação como algo essencialmente distinto de computação, tal como esta é tradicionalmente concebida. A ilustração mais detalhada da perspectiva de Varela, Thompson e Rosch está na discussão que eles fazem sobre experiência e categorização de cores, uma discussão que recebeu muita atenção em outros locais (*vide* THOMPSON; PALACIOS; VARELA, 1992; THOMPSON, 1995), normalmente sem referência às alegações mais abrangentes sobre corporificação, fenomenologia e Budismo feitas em *The Embodied Mind* (*vide* THOMPSON, 2007).

Desde a sua origem, a tradição enativista cresceu e enriqueceu de vários modos e cada uma de suas vertentes, ainda que compartilhando um quadro de referências comum com explicações vizinhas na ciência cognitiva corporificada, desenvolveu o seu próprio traço teórico e explorou um território conceitual que diferencia as vertentes entre si. Uma vertente, rotulada enativismo autopoietico, desenvolveu, em particular, o fenômeno biológico da *autopoiesis* e atribui a ele uma função central na explicação das propriedades fundamentais de nossa vida mental (MATURANA; VARELA, 1992; THOMPSON, 2005; DI PAOLO; THOMPSON, 2014). A noção de *autopoiesis* descreve sistemas vivos como ativos, adaptativos, autopreservativos e autoindividuaes, isto é, como tendo a propriedade de se autorreproduzir através de estratégias autorregulatórias. Outra vertente, o assim chamado enativismo sensório-motor, tomou uma rota mais liberal que, ao invés de sublinhar a função da organização autopoietica em sistemas vivos, busca extrair a compreensão implícita da dependência sensório-motora (isto é, o modo pelo qual o estímulo altera de modo contingente as ações dos organismos) para explicar a experiência consciente e um amplo grupo de comportamentos (NOE, 2004; O'REGAN, 2011; DEGENAAR; O'REGAN, 2017). Ainda outra vertente foi tão radical a ponto de afirmar que a cognição e a experiência não dependem da organização autopoietica e que a explicação sensório-motora permanece comprometida com o mentalismo e com o pensamento representacional (HUTTO; MYIN, 2013). Em suma, embora essas respectivas vertentes sejam amplamente consoantes, elas enfatizam diferentes aspectos, e as tensões internas motivaram uma larga fragmentação, várias formas de elaboração e prolongamento. Apesar de nos referirmos a outras vertentes que fazem um discurso definitivo sobre a linha de demarcação entre variedades de propostas enativistas (DEGENAAR; O'REAN, 2017; MENARY, 2006; HUTTO; MYIN, 2013), aqui nos limitaremos a uma visão geral panorâmica sobre algumas razões de suas discordâncias. Uma dessas razões está relacionada ao papel atribuído à noção de *autopoiesis*. Enquanto o programa alinhado a Varela, Thompson e Rosch sustenta que a *autopoiesis* (ou autoprodução) é uma pré-condição necessária da experiência, o que leva à ideia de que há uma profunda continuidade entre vida e mente (THOMPSON, 2007) e que a consciência é uma forma particular de vida, tanto o enativismo sensório-motor quanto sua formulação radical não colocam o processo autopoietico no centro e não assumem a visão de que o fenômeno cognitivo é construído sobre aspectos peculiares da organização de organismos vivos (DEGENAAR; O'REGAN, 2017; HUTTO; MYIN, 2013). Uma razão posterior para discordância diz respeito a apresentação da reivindicação de que a experiência

perceptual exige mediação do saber-como. Enquanto o enativismo sensório-motor sustenta que a experiência perceptual se faz possível pela posse e pelo exercício habilidoso do conhecimento prático das contingências sensório-motoras, as outras vertentes consideram o discurso sobre conhecimento mediato como uma marca do cognitivismo (HUTTO; MYIN, 2013), um tanto consistente com a teoria da mente que fundamenta a ciência cognitiva tradicional. Em sua expressão radical, o pensamento enativo sustenta que a mentalidade deve ser explicada em termos de engajamentos corporificados diretos que envolvem o ambiente.

### 2.3 Repensando a Robótica

No início dos anos 90, o trabalho em inteligência computacional tinha começado a explorar formas de gerar ação inteligente em robôs que logo ficou conhecida como a abordagem corporificada da robótica. Em um conjunto de artigos, Rodney Brooks (1991a, 1991b) apresentou um panorama geral e acessível de um novo tipo de arquitetura computacional inteligente, uma arquitetura de subsunção, que era de leve representação e dirigida ao mundo. Nesses aspectos, essa arquitetura se afasta de visões tradicionais de planejamento e tomada de decisão que, assentadas em intenso processamento de representações, caracterizou a clássica IA, e foi caracterizada por Brooks como fornecendo “inteligência sem representação”. Junto com o trabalho computacional de Agre e Chapman (1987) e Suchman (1987), a abordagem de Brooks sugere uma concepção da inteligência computacional na qual o controle era governado de baixo para cima [*bottom-up*] pelo comportamento e pela interação com o mundo, e não por abundantes e frequentemente complicados algoritmos e representações internas.

O alcance do trabalho em robótica reativa, ou baseada no comportamento, e sua identificação como marcando uma parte da ciência cognitiva corporificada, foi anunciada por Andy Clark em *Being There: Putting Mind, World, and Body Back Together* (1997). Aqui, Clark ofereceu um quadro de referências integrado em uma ampla gama de trabalhos que emergem sobre corporificação nas ciências cognitivas. A grande ideia em *Being There*, com um impacto duradouro na ciência cognitiva corporificada, é que mentes não são para pensar, como tradicionalmente concebidas, **mas para fazer**, para fazer as coisas no mundo em tempo real. Ao invés de desenvolver “enciclopédias ambulantes”, a robótica no fim dos anos 80 e início dos anos 90 começava a se concentrar na interação dinâmica entre corpo e mundo.

Clark esboçou afinidades entre essa mudança na concepção da ação inteligente em sistemas computacionais e a emergência da ideia de que a cognição era escalonada, situada e estendida.

Os trabalhos que brevemente descrevemos até aqui dizem respeito ao que chamamos de cognição no sentido estrito, a processos como a memória humana, categorização, processamento da linguagem (Lakoff e Johnson), categorização de cor por humanos e não-humanos (Varela, Thompson e Rosch) e tomadas de decisão e planejamento em robôs e sistemas de robótica (Clark). Mas a ciência cognitiva corporificada visa abranger a cognição amplamente construída. A fim de aqui transmitir o sabor de trabalhos iniciais, consideremos brevemente a obra de James Gibson sobre a percepção, e o trabalho de Esther Thelen e Linda Smith (1994) sobre o caminhar e o comportamento de alcançar coisas das crianças (*vide* SHAPIRO, 2011).

## 2.4 Percepção Ecológica

A explicação da visão proposta por James Gibson (1979) desafiou a ideia de que o problema central que o sistema visual tem para solucionar é como reconstruir um mundo desenvolvido e tridimensional a partir da informação especificada na imagem bidimensional da retina. Esta ideia foi proeminente na concepção tradicional da visão, de processamento da informação, a qual inclui Rock (1983, 1997), Richard Gregory (1966) e Marr (1982). Gibson pensa que esse não é um problema que o sistema visual enfrenta porque a visão não começa com uma matriz estática da retina, mas com um organismo que ativamente se move através de um ambiente visualmente rico. A abordagem positiva de Gibson da visão tentava especificar essa riqueza, a informação naquilo que ele chamava de *matriz óptica do ambiente*, especialmente invariantes nesta matriz, os quais podem ser usados para distinguir aspectos que dependem do agente e os aspectos objetivos do ambiente de alguém. Por enfatizar tanto a função do movimento daquele que percebe quanto a integração deste em um ambiente maior e visualmente rico, Gibson é considerado como ao menos um proponente incipiente da concepção corporificada (*vide* WILSON, 2004, cap. 7; SHAPIRO, 2011, cap. 2).

## 2.5 Dinamismo e Desenvolvimento

Esther Thelen e Linda Smith (1994) ofereceram um desafio radical às concepções tradicionais nativistas do desenvolvimento cognitivo ao aplicar a teoria dos sistemas dinâmicos à psicologia do desenvolvimento. Uma importante implicação da teoria dos sistemas dinâmicos é que os sistemas podem gerar novos comportamentos, por exemplo, diferentes soluções para alcançar objetos, através da atividade corpórea, questionando a necessidade de postular padrões pré-programados que se desdobram com o tempo. Ao levantar questões fundamentais sobre suposições compartilhadas na área, elas argumentam que o comportamento de caminhar em crianças não é dirigido por processos de amadurecimento que são determinados de algum modo por um fio rígido do código genético, mas resulta, ao invés disso, da interação entre os movimentos iniciais espontâneos dos membros das crianças e as mudanças de contextos. Assim, elas observaram esse aspecto particular do desenvolvimento como um produto emergente e auto-organizador de muitas interações descentralizadas e locais acontecendo em tempo real, com a promessa de generalizar essa abordagem para o desenvolvimento cognitivo de modo mais geral.

## 2.6 Fenomenologia

Finalmente, a fim de ancorar a história recente, a ideia de que uma compreensão do corpo sustenta a possibilidade mesma **da experiência** tem raízes nas obras fenomenológicas de Edmund Husserl (1913, 1931), Maurice Merleau-Ponty (1945) e Jean-Paul Sartre (1943), raízes que vimos reconhecidas por Varela, Thompson e Rosch em *The Embodied Mind*. Essa tradição continental mais antiga foi explorada construtivamente desde o início na inteligência artificial, com especial menção a Heidegger, por Winograd e Flores (1986), e também formou o pano de fundo para a crítica clássica de Dreyfus (1972) ao computacionalismo tradicional.

A ciência cognitiva corporificada impele as explicações fenomenológicas em novas direções. Ela busca não tanto compreender como a fisicalidade torna acessível a experiência do eu [*self*], do mundo e dos outros, mas, antes, pretende especificar os **mecanismos** que explicam precisamente como a cognição está baseada na, e profundamente condicionada pela, natureza corpórea da agência cognitiva. Não iremos explorar a convergência entre a tradição fenomenológica antiga e a ciência cognitiva corporificada, embora reconheçamos que os *insights*

fenomenológicos podem ser uma fonte indispensável para a investigação em andamento da consciência, autoconsciência, ação e intersubjetividade (ver GALLAGHER, 2009; GALLAGHER; ZAHAVI, 2008; THOMPSON, 2007; GALLAGHER, 2005; WHEELER, 2005).

### 3. O que a Cognição Corporificada é

A caracterização geral da cognição corporificada com a qual começamos providencia a base para o que chamaremos a Tese da Corporificação:

**Tese da Corporificação:** Muitos aspectos da cognição são corporificados no sentido de que são profundamente dependentes de características do corpo físico de um agente, tal que o corpo além-do-cérebro do agente desempenha uma função causal significativa, ou um papel fisicamente constitutivo, no processamento cognitivo desse agente.

Todos os cinco exemplos anteriores de trabalhos em cognição corporificada que brevemente resumimos na Seção 2 aceitam a Tese da Corporificação. A hipótese de trabalho da ciência cognitiva corporificada é a de que essa tese é verdadeira, quer seja pelo significativo papel causal, quer seja pelo significativo papel fisicamente constitutivo do corpo no processamento cognitivo. Os proponentes da ciência cognitiva corporificada defendem tanto a alegação causal quanto a constitutiva sobre a função do corpo na cognição. Embora a atribuição de uma função fisicamente constitutiva ao corpo na cognição tenha sido tomada como desafiando a ciência cognitiva tradicional de um modo mais radical do que o da mera função causal significativa, ambas as versões da Tese da Corporificação marcam um afastamento das concepções da mente dominantes na ciência cognitiva tradicional.

Ao invés de seguir aqueles que tentam explicar a cognição corporificada por apelo à metáfora do enraizamento [*grounding*] (vide ANDERSON, 2003; BARSALOU, 1999, 2008; GLENBERG; ROBERTSON, 2000; GLENBERG *et al.*, 2005), nós consideramos que a melhor maneira de articular ainda mais a Tese da Corporificação é especificar a natureza da dependência da cognição sobre o corpo: qual particular papel causal significado ou qual a função fisicamente constitutiva o



corpo desempenha na cognição? (*vide* SHAPIRO, 2010, 2011; CLARK, 2008; THOMPSON, 2007; WHEELER, 2005; ANDERSON, 2003; WILSON, 2002).

No nível mais geral, existem três distintas funções ou papéis, cada um deles com suas próprias implicações para a ciência cognitiva corporificada. O corpo pode funcionar como uma **restrição** sobre a cognição, como um **distribuidor** para o processamento cognitivo e como um **regulador** da atividade cognitiva.

Podemos especificar cada uma dessas funções ou papéis de modo mais preciso, e chamar a atenção para as consequências distintivas que elas possuem e para o trabalho já descrito que apela a cada uma dessas concepções da Tese da Corporificação.

**Corpo como Restrição:** o corpo de um agente funciona para restringir significativamente a natureza e o conteúdo das representações processadas pelo sistema cognitivo desse agente.

Dentre as alegadas consequências da tese do Corpo como Restrição existem duas que gostaríamos de chamar a atenção:

- Algumas formas de cognição serão mais fáceis e virão mais naturalmente devido às características corpóreas de um agente; do mesmo modo, alguns tipos de cognição serão mais difíceis ou ainda impossíveis devido ao corpo que um agente cognitivo tem.
- A variação cognitiva é às vezes explicada por um apelo à variação corpórea.

O trabalho de Lakoff e Johnson sobre impregnação/permeação [*permeation*] da cognição pela metáfora, e o de Varela, Thompson e Rosch sobre a cognição enativa (especialmente no domínio da percepção e categorização de cores), exemplificam ambos a tese do Corpo como Restrição.

**Corpo como Distribuidor:** o corpo de um agente funciona para distribuir cargas computacionais e representacionais entre estruturas neurais e não-neurais.

Diferente da função do corpo na tese do Corpo como Restrição, aqui o corpo é tomado como compartilhando seu próprio processamento cognitivo, servindo para distribuir tarefas cognitivas entre cérebro e corpo. A tese do Corpo como Distribuidor tem três implicações importantes que merecem serem explicitadas:

- Estruturas cognitivas neurais (*neural-realized*) podem ser mais raras do que foi tradicionalmente assumido e, em princípio, totalmente ausentes.
- Estruturas corpóreas podem por si mesmas serem ao menos realizadoras parciais do maquinário físico que realiza processos cognitivos.
- A cognição não é limitada pelo crânio, assim, sistemas cognitivos podem incluir tanto partes não-neurais do corpo, quanto o ambiente além-do-corpo.

Como essas implicações deveriam sugerir, é o Corpo como Distribuidor que está em jogo para aqueles que usam a cognição corporificada para desafiar as concepções tradicionais da representação mental (GIBSON, 1979; THELEN; SMITH, 1994; *vide* GLENBERG, 1997, sobre memória; SHAPIRO, 1997; WILSON, 2004: cap. 7-8, sobre representação exploradora). O apelo à *computação morfológica* (MACIVER, 2009), segundo a qual propriedades de estruturas anatômicas (tal como a forma das orelhas de um morcego) desempenham uma função computacional em um processo cognitivo (tal como a ecolocalização), também depende da tese do Corpo como Distribuidor. E ao subsumir as estruturas corpóreas de um agente, assim como os aspectos do ambiente como formas de estruturas não-neurais, a tese do Corpo como Distribuidor traça uma conexão entre a cognição corporificada e versões da tese da mente estendida que apela a conceitos como realização e escalonamento (WILSON; CLARK, 2009; WILSON, 2004: cap. 5-6; CLARK, 2003).

Intimamente relacionada a tese do Corpo como Distribuidor está:

**Corpo como Regulador:** o corpo de um agente funciona para regular atividades cognitivas no espaço e no tempo, garantindo que a cognição e a ação sejam fortemente coordenadas.

Distinguimos essa versão da Tese da Corporificação da tese do Corpo como Distribuidor por conta das supostas implicações distintivas que atribuir uma

função regulativa ao corpo têm na cognição. Essas incluem:

- Estruturas corpóreas facilitam a execução em tempo real dos comportamentos complexos como reação aos eventos ambientais complexos e mutáveis.
- O corpo não meramente funciona para a conversão [*transduce*] das entradas do mundo [*world-inputs*] à cognição e depois entregar as saídas ao mundo [*worldly-outputs*] na forma do comportamento a partir de processamentos cognitivos internos, mas é integral ao controle online da própria cognição.

Aqui o corpo tem uma função dirigida pela reação no processamento cognitivo e a tese do Corpo como Regulador foi especialmente proeminente nas abordagens dinâmicas da cognição (*vide* PORT; VAN GELDER, 1995; BEER, 2000; THELEN; SMITH, 1994; CEMERO, 2009).

Para resumir essa seção: distinguimos três modos de articular a Tese da Cognição Corporificada, cada uma das quais especifica uma maneira particular pela qual a cognição depende do corpo. Para colocar de modo mais categórico (e pensamos, informativamente), existem três funções ou papéis distintivos para o corpo que a ciência cognitiva corporificada poderia atribuir: como uma restrição sobre a cognição, como um distribuidor para o processamento cognitivo e como um regulador em tempo real da atividade cognitiva. Essas determinadas formas da Tese da Corporificação podem atribuir ao corpo ou uma função causal significativa, ou uma função fisicamente constitutiva, na cognição.

#### **4. Corporificação vs. Tradição sobre Três Questões**

Nessa seção, exploramos a promessa revolucionária da ciência cognitiva corporificada com respeito a três tópicos comuns na filosofia da mente e na ciência cognitiva: a modularidade da mente, a natureza da representação mental e o nativismo. Para posições alternativas da cognição situada na modularidade, representação e nativismo, confira Bechtel (2009), Rowlands (2009), Rupert (2009a). Começemos com alguns contrastes importantes gerais entre a ciência cognitiva tradicional e a ciência cognitiva corporificada.

As concepções tradicionais tendem a assumir a existência de representações discretas e internas realizadas por mecanismos subjacentes, nitidamente distintos e altamente especificados no cérebro. Estes mecanismos, por sua vez, foram moldados pela seleção natural e codificados em estruturas genéticas. Assim, as posições tradicionais têm influenciado na neurociência e se comprometido com o individualismo ou o internalismo, a defesa de que a cognição sobrevém as propriedades intrínsecas e físicas do sujeito cognoscente. A estratégia de pesquisa do “solipsismo metodológico” (FODOR, 1980, 1981) é uma versão clássica dessa concepção individualista da cognição. O modo pelo qual tópicos centrais foram abordados reflete profundamente a ideia de que fenômenos cognitivos podem ser explicados localmente e que os elementos além dos limites do crânio são de interesse apenas na medida em que eles oferecem entradas sensoriais e permitem saídas comportamentais. Nas palavras de Susan Hurley (1998), as posições principais sobre a mente se comprometeram com o “modelo clássico do sanduíche”, a afirmação de que a cognição (no sentido estrito) é segregada do processamento de sistemas de baixo nível, assim agindo como carne em um sanduíche empanado pela percepção e pela ação.

A ciência cognitiva corporificada, por contraste, modelou a cognição como o produto da interação dinâmica entre processos neurais e não-neurais, sem qualquer fratura geral entre cognição, experiência corpórea do agente e contextos da vida real. Aqui o corpo é considerado como restringindo, distribuindo ou regulando o processamento cognitivo. Especificar como o corpo realiza essas funções em ambientes particulares levanta o prospecto de que a própria cognição não é limitada pelo cérebro, nem talvez ainda pelo próprio corpo.

Podemos expressar o contraste *prima facie* entre ciência cognitiva tradicional e corporificada em termos das posições opostas sobre representação mental, computação e realização. Enquanto muitas explicações tradicionais tendem a ver a cognição como representacionalmente localista, computacionalmente fixa e adequadamente caracterizada sem depender do sistema neural que a realiza, dos aspectos do corpo físico e do ambiente circundante, a ciência cognitiva corporificada tende a ver a cognição como representacionalmente distribuída, computacionalmente dinâmica e como adequadamente caracterizada apenas por referência aos detalhes da realização corpórea. Desse modo, a ciência cognitiva corporificada motiva uma interrogação de algumas suposições fundamentais feitas na ciência cognitiva. Tal exame é também manifesto em posições de modelos cognitivos em que eles próprios não são especialmente corporificados, tal como o conexionismo, que deixamos de lado aqui.

## 4.1 Modularidade

Sistemas modulares são independentes, de domínio específico, encapsulados, conectados [*hardwired*] e funcionam em um processamento hierárquico de baixo para cima. Quando a teoria da modularidade foi introduzida na ciência cognitiva (FODOR, 1983), a cognição central – cognição no sentido estrito – foi caracterizada como não-modular e como claramente distinta dos sistemas modulares periféricos, tais como aqueles que governam a percepção e o controle motor (acrescentando, como Fodor diz, a linguagem). As afirmações de Fodor sobre a cognição central foram colocadas em questão por uma grande variedade de pesquisadores (*vide* CARRUTHERS, 2006; SPERBER, 2001; COSMIDES; TOOBY, 1997; HIRSHFELD; GELMAN, 1994; WILSON, 2005, 2008). Esses críticos argumentam que os processos cognitivos mais altos também atendem aos critérios de modularidade. De fato, a posição de que a cognição no sentido estrito é também modular é suficientemente difundida pela psicologia do desenvolvimento, evolucionária e cognitiva, talvez se colocando como a forma dominante da teoria da modularidade na ciência cognitiva contemporânea, a despeito da posição contrária de Fodor (2000).

Tanto a explicação tradicional, a de Fodor sobre modularidade, quanto as de seus críticos e descendentes dominantes, fazem referência ao corpo e ao ambiente apenas indiretamente. A teoria da modularidade, em qualquer uma de suas versões, tem pouca coisa positiva a dizer sobre a carga real das entidades que estão além do crânio na performance cognitiva, apenas vendo elas como fontes de entrada e repositórios de saídas comportamentais. Essa posição implica que todo o trabalho duro, cognitivamente falando, é realizado pela cognição central sozinha, com processos periféricos simplesmente oferecendo entradas e executando instruções. As afirmações de que o processamento cognitivo ocorre puramente no cérebro num estilo modular e cumpre sua tarefa ao operar essencialmente independentemente do planejamento e da execução motora, entretanto, são colocadas em questão por estudos empíricos da experiência corporificada.

Um exemplo de experiência corporificada em relação à linguagem, que exemplifica a tese do Corpo como Restrição, é a Hipótese Indexical de Glenberg (GLENBERG *et al.*, 2009; GLENBERG; KASCHAK, 2002; GLENBERG; ROBERTSON, 1999, 2000). Essa concepção defende que a compreensão de uma frase é realizada através da ativação de esquemas relevantes de ação, recrutando os mesmos

mecanismos neurais ativos no comportamento manifesto e pela combinação de *affordances*. Considere as frases:

(1a) Depois de caminhar descalço pelo lago, Erik usou sua camisa para secar seus pés.

(1b) Depois de caminhar descalço pelo lago, Erik usou seus óculos para secar seus pés.

Embora tanto (1a) quanto (1b) sejam gramaticalmente corretas, a razão pela qual (1b) não faz o mesmo tipo de sentido que (1a) é que as *affordances* dos óculos não combinam com a ação de secar. Compreender o significado de tais frases exige conhecer as possibilidades oferecidas pelos objetos referidos nelas. Essas possibilidades são condicionadas pela interação entre as capacidades corpóreas e os referentes. Outro estudo mostrou que pessoas são mais rápidas para compreender frases onde objetos oferecem a *affordance* necessária para realizar uma ação, por exemplo, cadeira com quatro rodas para mover caixas grandes, do que frases nas quais objetos não oferecem *affordances*, por exemplo, uma cadeira com quatro rodas faltando (KASCHAK; GLENBERG, 2000).

Tais descobertas indicam que a construção do significado é condicionada pelas possibilidades corporificadas que um cenário oferece, e sugerem que os processos sensório-motores contribuem para a compreensão da linguagem. Essa conclusão seria provavelmente rejeitada por defensores da teoria modular, pois os seus compromissos com o encapsulamento e o domínio específico implicam que o processamento da linguagem não pode ser modulado por informação motora e pelo conhecimento de fundo. Borghi, Glenberg e Kaschak (2004) também relataram que na compreensão da linguagem a perspectiva implicada pela frase orienta a recuperação da informação sobre objetos, tornando disponível o conhecimento conceitual. Respostas a questões sobre se um objeto (uma mesa, digamos) é parte da situação descrita na frase são mais rápidas se há compatibilidade entre o nome do objeto e a perspectiva implicada pela frase (comer em um restaurante, por exemplo). Respostas corporificadas são ativadas também ao julgar propriedades específicas de objetos, sugerindo que a visão e a ação são intimamente integradas no organismo biológico e que elas juntas condicionam o processamento cognitivo.

A nítida distinção entre visão e ação que é parte da explicação modular da cognição também foi desafiada por estudos da experiência corporificada. Por exemplo, quando as pessoas são solicitadas a escolher, entre escadas de diferentes

alturas, aquela em que elas podem subir mais facilmente, suas respostas são consistentes com respeito às suas habilidades de subir escadas (WARREN, 1984). Resultados similares foram descritos por juízos de agarrar objetos (VAN LEEUWEN *et al.*, 1994), pegar bolas (OUDEJAMS *et al.*, 1996) e escalar paredes (WAGMAN; CARELLO, 2001). Estudos como esses apoiam a afirmação geral de que a experiência perceptual incorpora a antecipada interação corporificada, sugerindo que visão e ação são integradas, ou ao menos interligadas, de modos que são incompatíveis com o modelo unidirecional da cognição postulado por modularistas (HURLEY, 1998; WILSON, 2010).

Até mesmo atribuir emoções, intenções e crenças a alguém parece pressupor uma certa realização corpórea. Diferente das posições tradicionais, que postulam um módulo inato de Teoria da Mente para explicar a cognição social (LESLIE, 1987; BARON-COHEN, 1995), um número crescente de estudos no campo (RIZZOLATTI; CRAIGHERO, 2004; RIZZOLATTI *et al.*, 1996; GALLESE *et al.*, 1996), sugere que a compreensão de outras mentes é baseada primariamente na expertise motora subjacente a nossa capacidade de agir. Tal compreensão corporificada não é apenas diferente por natureza das modalidades de leitura da mente [*mindreading*] como tradicionalmente concebidas, mas também indica fortemente que o significado do comportamento intencional pode ser apreendido somente se conhecemos corporalmente, ou experiencialmente, ou ambos, como é estar em um estado mental [*what it is like to be in a mental state*].

O processo que subscreve uma variedade de fenômenos mentais – discernir as frases significativas daquelas que não são, extrair as possibilidades proporcionadas pelos objetos e detectar comportamentos intencionais – exige a contribuição orquestrada de muitos componentes, neurais e não-neurais. Se estados corpóreos e sistemas de modalidades específicas do cérebro servem como o fundamento de vários aspectos de nossa vida cognitiva, então os módulos de domínios específicos tradicionais não são os elementos significativos de análise que eles pretendiam ser sob a Tese da Descorporificação. A afirmação de que a cognição depende profundamente dos processos que evoluíram para permitir aos organismos interagir efetivamente com o ambiente sugere que a mente não está desacoplada da experiência corporificada da maneira pressuposta pelas posições tradicionais na ciência cognitiva. Antes, o corpo pode agir como uma **restrição** da cognição e como um **distribuidor** do processo cognitivo (*vide* Seção 3).

## 4.2 Representação Mental

No coração da compreensão científica tradicional da cognição encontra-se uma concepção particular de representação mental (FODOR; PYLYSHYN 1988; NEWELL; SIMON, 1972). Essa concepção sustenta que representações são estruturas simbólicas com propriedades semi-linguísticas [*quasi-linguistic*] e combinatórias, que atuam como veículos de conteúdos e são o principal recurso para explicar o comportamento inteligente. Representações mentais são simbólicas e abstratas na medida em que a mesma representação, tal como “mesa”, é usada para significar diferentes tipos de mesas. Elas são amodais no sentido em que a mesma representação pode ser empregada quando se escreve ou fala de “mesa”. Tais representações são arbitrariamente relacionadas aos seus referentes, pois o modo pelo qual elas são formadas e implantadas, junto com suas características, não têm qualquer relacionamento com os aspectos físicos e funcionais dos referentes. Assim, segundo a posição tradicional, representações internas empregues na linguagem, na formação de conceitos e na memória não são apenas essencialmente distintas daquelas [representações] processadas pelo sistema sensorio motor, mas seu significado é divorciado da experiência corpórea. Defensores dessa posição se comprometem com ao menos três princípios fundamentais.

- A informação transmitida por uma representação mental não exhibe nenhum aspecto específico de modalidade. Nesse sentido, representações são autônomas em relação aos sistemas perceptuais, a ação corpórea e aos seus detalhes operacionais;
- O conhecimento é organizado proposicionalmente, com o significado das palavras emergindo de suas relações com símbolos internos. Determinar o significado de um símbolo é como procurar no dicionário a fim de encontrar qual definição é dada por suas relações com outros símbolos;
- Representações internas costumam ser usadas para instruir programas motores, os quais são essencialmente separados e independentes da cognição. Assim, o processamento cognitivo não é inextrincavelmente moldado por ações corpóreas.



Recentemente, diversas explicações alternativas para o comportamento adaptativo ganharam atenção. Embora diversificadas, todas elas colocaram em questão o compromisso com esses princípios. Ao defender a ideia de que a atividade cognitiva reutiliza os processos e as representações implantadas na percepção e na ação, essas explicações colocam um sério desafio à noção de que as representações conceituais e semânticas são (ou devem ser) amodais. Permanecem diferenças significativas entre essas posições com respeito à radicalidade de suas inclinações antirrepresentacionistas (*vide* CHEMERO, 2009; HUTTO; MYIN, 2013; MYIN; DEGENAAR, 2014).

Entre as posições antirrepresentacionistas mais influentes está a teoria dos sistemas dinâmicos (BEER, 1990, 2002, 2003; BROOKS, 1991a, 1991b, 2002; THELEN; SMITH, 1994; VAN GELDER, 1992). Dinamicistas tendem a minimizar e às vezes até mesmo negar a necessidade de um processador representacional centralizado. A noção de representação que esses autores desafiam é relativamente específica: um modelo interno capaz de reproduzir a estrutura do ambiente externo que é usado pelo agente cognitivo para orientar seu comportamento com certa independência do mundo. A teoria dos sistemas dinâmicos mostra ser popular na robótica e no trabalho sobre vida artificial, em que se tenta explicar o comportamento adaptativo em termos de corporificação e situacidade [*embeddedness*]. Enquanto uma criatura situada puder sentir seu mundo de modo a permitir que seu corpo seja diretamente influenciado, descrições simbólicas abstratas podem ser dispensadas.

Formular uma teoria empiricamente adequada do comportamento inteligente sem apelar a quaisquer representações, no entanto, enfrenta dificuldades insuperáveis, e a ideia de que é uma questão relativamente trivial escalar a partir dos modelos dinâmicos existentes para explicar toda a cognição permanece uma quimera e suscetível aos mesmos problemas que motivaram a mudança do comportamentalismo à ciência cognitiva em primeiro lugar. Por exemplo, a interação organismo-ambiente não pode por si mesma explicar o comportamento antecipatório, o qual envolve fatores internos além das restrições imediatas do ambiente para executar ou satisfazer futuras necessidades, objetivos e condições. Os campos que levantam um problema de fome de representação [*representation-hungry*] (CLARK, 1997) são aqueles que envolvem raciocinar, planejar, imaginar e interagir sobre um estado de coisas ausente, não-existente ou contrafactual. Para uma posição recente contrastante, consulte Chemero (2009).

Além do mais, não é claro porque a ciência cognitiva corporificada não poderia também ser simbólica, representacional, abstrata, etc. A perplexidade aqui

é ampliada pelo fato de que muitas abordagens corporificadas da cognição, com seus estilos próprios, são simbólicas, representacionais, abstratas, etc. O que elas oferecem são concepções específicas do que os símbolos e representações mentais são, e como funcionam em nossa economia cognitiva. Tipicamente, elas substituem a concepção de codificação proposicional por uma de acordo com a qual os símbolos transmitem um aspecto específico da modalidade.

Uma concepção que adapta a noção de representação mental, ao invés de dispensá-la, é a teoria dos símbolos perceptuais de Lawrence Barsalou (1999, 2003, 2008, 2009). Essa teoria repousa sobre a suposição de que a cognição humana não consiste em representações amodais que mantêm relações arbitrárias com os seus referentes no mundo, mas, antes, representações cujos padrões de ativação incluem informação de várias modalidades sensoriais. Por exemplo, a estrutura simbólica que representa a cor de um objeto em sua ausência, digamos, durante a visualização mental ou em tarefas de solução de problemas, dependem do mesmo sistema neural que é recrutado quando a cor é efetivamente percebida. Assim, não apenas os mecanismos cognitivos e perceptuais compartilham estados representacionais, mas o processamento cognitivo essencialmente reativa áreas sensorio-motoras para executar simulações perceptuais. Uma implicação posterior é que símbolos perceptuais não são independentes do sistema biológico que os incorpora e o conteúdo transmitido provavelmente variaria conforme os sistemas inteligentes variassem fisicamente. É devido a esse apelo ao corpo e aos processos sensorio-motores que condicionam a natureza dos símbolos disponíveis à cognição que a teoria de Barsalou exemplifica a tese do Corpo como Restrição.

Enquanto defensores das representações modais argumentam que há pouca evidência empírica direta para as representações amodais, pela adoção do modelo tradicional da representação amplamente motivado por razões teóricas (BARSALOU *et al.*, 2003, 85), que suas próprias posições são empiricamente plausíveis (*vide* PECHER *et al.*, 2003; ZWAAN; YAXLEY, 2003; GLENBERG; KASCHAK, 2002; MARTIN; CHAO, 2001; SOLOMON; BARSALOU, 2001; MARTIN *et al.*, 2000; SPIVEY *et al.*, 2000), eles também subestimam a dificuldade de oferecer evidência definitiva contra explicações amodais. Como Machery (2007) aponta, as performances dos sujeitos podem ser acomodadas por explicações modais e amodais (*vide* RUPERT, 2006). Posto que a maioria dos defensores das explicações amodais (*vide* SIMON, 1995; FODOR, 1975) não nega o envolvimento de áreas perceptuais durante as tarefas cognitivas, tal como a imagem visual, mostrar que o córtex visual é ativado quando, digamos, se imagina o verde de uma maçã não

fornece evidência para a teoria de Barsalou em relação à explicação amodal. Além disso, como tradicionalistas apontam, alguns problemas mentais são solucionados sem imagens (relatáveis) e indivíduos às vezes recorrem ao conhecimento arquivado em sistemas representacionais livres de modalidade. Consequentemente, generalizar resultados sobre o uso de estimulações perceptuais de algumas para todas as tarefas, o que é exigido para uma teoria modal geral da representação, é algo problemático.

### 4.3 Nativismo

Atualmente, muitos pesquisadores reconhecem que a cognição se desenvolve como um resultado da interação dinâmica entre o inato e o aprendido, ainda que não seja sempre claro como essa interação se dá. Segundo a posição bidimensional do nativismo defendida por um de nós em outro lugar, que distingue entre formas fortes e fracas do nativismo (WILSON, 2004, cap.3), nativistas fortes se comprometem com as duas teses seguintes:

- 1 As estruturas e os processos internos necessários para o desenvolvimento de um indivíduo são ricos, complexos e causalmente poderosos;
- 2 Processos externos ao indivíduo desempenham uma função causal secundária na aquisição e desenvolvimento dessas estruturas.
- 3 Antinativistas fortes (tal como os empiristas clássicos), por contraste, negam essas concepções gerais.

Os paradigmas comprometidos com o nativismo forte produziram resultados notáveis no campo do desenvolvimento cognitivo de crianças, incluindo áreas como aritmética e física (BAILLARGEON, 2002; BAILLARGEON *et al.*, 1985; SPELKE *et al.*, 1992, 1995). Os críticos do nativismo rotularam o nativismo como antidesenvolvimentista por natureza (HAITH, 1998) e alegaram que as conclusões de que a mente consiste em traços interligados que se desdobram por maturação conforme um padrão fixo conflitam com a própria ideia de aprendizagem e flexibilidade (QUARTZ; SEJNOWSKI, 1997; THELEN; SMITH, 1994; BATES *et al.*, 1995; KARMILOFF-SMITH, 1994). Se o nativismo fosse verdadeiro, então o corpo e o ambiente (incluindo a cultura), mesmo relevantes, deveriam ser considerados como

meros “gatilhos” de aspectos ontogeneticamente determinados que se desenvolvem de maneiras previsíveis.

Embora à primeira vista possa ser pensado que a cognição corporificada não tem implicações distintivas para a discussão corrente entre “nativismo” e “empirismo”, uma contribuição da ciência cognitiva corporificada aqui está em sua exploração específica das funções que o corpo desempenha no processamento cognitivo. Essas funções frequentemente colocam desafios a posições como a do nativista forte e a do empirista forte. Como tal, a ciência cognitiva corporificada não simplesmente assume, com os empiristas, que o processamento cognitivo depende em grande medida da exposição ao ambiente e que a cognição é uma reflexão causal dele. Além disso, enquanto empiristas normalmente concebem o mundo como alguma coisa objetivamente dada ao sujeito, que assim forma uma representação estática dele que então guia sua ação, a ciência cognitiva corporificada aborda como a interação dinâmica entre o agente corporificado e o mundo gera cognição. É esse foco sobre a interação dinâmica e mundana que oferece uma conexão da cognição corporificada à situada dentro da ciência cognitiva situada. Aqui vamos explorar se, e em que medida, a ciência cognitiva corporificada solapa as afirmações do nativista forte.

De modo amplo, a ciência cognitiva corporificada geralmente minimiza a riqueza interna necessária para realizar o comportamento complexo (RUPERT, 2009a), enfatizando o fato de que o processamento e o desenvolvimento cognitivo são profundamente afetados pela perturbação através das mudanças no ambiente e na ação do corpo (GRIFFITHS; STOTZ, 2000; CLARK; TORIBIO, 1994; THELEN; SMITH, 1994; VARELA; THOMPSON; ROSCH, 1991; BALLARD, 1991). O corpo em ação é uma restrição poderosa sobre como organismos concebem seus nichos, pois essa restrição permite que certas interações e experiências tenham um efeito na formação de conceitos e na compreensão do significado linguístico. Por exemplo, por ter dedos capazes de agarrar objetos e pernas capazes de caminhar e escalar paredes, classificamos e categorizamos estímulos de modos que são radicalmente diferentes de, digamos, os modos em que são classificados por borboletas. A própria experiência sinestésica ordinária essencialmente enquadra a aquisição e o desenvolvimento de estruturas cognitivas. Isto parece apoiar a contribuição significativa dos componentes além-do-crânio na realização do fenômeno cognitivo e, em termos do pano de fundo que introduzimos (*vide* seção 3), exemplifica tanto a tese do Corpo como Restrição quanto a tese do Corpo como Distribuidor.

Exemplos que apoiam a tese do Corpo como *Distribuidor* surgem da pesquisa não-nativista sobre a percepção. Muitos módulos presumidos, tal como aqueles para o processamento sensorio, se desenvolvem parcialmente via aprendizagem. No recém-nascido, as modalidades sensorias e vias corticais não são tão diferenciadas quanto as de cérebros maduros e parecem emergir através de uma série de interações fortalecedoras entre o corpo ativo e o ambiente (QUARTZ; SEJNOWSKI, 1997). Isso implica que em bebês há poucas, se é que há, vias corticais de domínio específico e altamente especializadas para a maioria das tarefas, mas durante o desenvolvimento em virtude de interações ativas com o ambiente são recrutadas e ajustadas para processamento de estímulos particulares (ELMAN *et al.*, 1996). Outros resultados empíricos no estudo de modalidades sensorias apontam na mesma direção, rejeitando a ideia que os sentidos são módulos que se conectam para a percepção no nascimento, com as interações do corpo com o ambiente desempenhando apenas uma função secundária em limitar ou mesmo determinar a natureza do processamento perceptual.

Por outro lado, parece que muitas capacidades relevantes não são tão específicas de um domínio quanto elas podem inicialmente parecer. Por exemplo, mesmo que o córtex visual pareça dedicado a processar uma classe particular de informação, ele pode ser recrutado por uma modalidade sensorial diferente durante a leitura do Braille – tenha o sujeito cegueira congênita, adquirida ou induzida (vendada) (SADATO *et al.*, 1996, 1998; PASCUAL-LEONE; HAMILTON, 2001; PASCUAL-LEONE *et al.*, 1998). Se estendermos essas observações ao reconhecimento facial, surge suporte adicional à hipótese de que a especialização começa em estágio posterior de desenvolvimento, em virtude da experiência. Apesar do fato de a área fusiforme do rosto [*fusiform face area*] (FFA) ser altamente seletiva para faces (KANWISHER, 2000; KANWISHER *et al.*, 1997), dados sugerem que ela também pode ser ativada em resposta a aspectos não-faciais, tais como pássaros e carros, desde que os sujeitos tenham uma expertise substancial nestas categorias (GAUTHIER *et al.*, 2000).

Ainda que essa evidência não coloque em questão a ideia da especialização cerebral, ela indica o papel da atividade corpórea na geração da diferenciação de vias corticais e na emergência de funções específicas, tais como as respostas seletivas de direção no córtex visual. Interações corporificadas com o mundo moldam e controlam os mecanismos responsáveis pelo processamento dessa informação, oferecendo suporte para as teses do Corpo como *Distribuidor* e do Corpo como *Regulador*. Além do mais, considere os estudos que investigam como trocar as

mãos para fazer algo molda as representações sensório motoras corticais dos movimentos dos dedos. Uma mudança da mão esquerda para a direita na escrita não apenas desencadeia uma reorganização geral da dominância motora, mas também tem um impacto amplo sobre a neuroanatomia funcional do sistema motor que controla as mãos, influenciando inclusive tarefas motoras que exigem pouca habilidade (KLOEPPEL *et al.*, 2007). Os padrões de reorganização encontrados em canhotos convertidos mostram quão flexível o cérebro é em termos de quais regiões podem fazer o que em resposta ao treinamento educacional, ao uso da mão durante a vida, e como aspectos corpóreos e esquemas de interação são condições às quais o cérebro está sintonizado. Estes estudos podem ser interpretados como sugerindo que interações corporificadas e o ambiente além-do-corpo podem por si mesmos serem realizadores parciais da atividade cognitiva correspondente. Incluir partes não-neurais do corpo entre os blocos de construção fisicamente constitutivos da cognição sugere uma leitura mais radical da cognição corporificada.

As afirmações de nativistas fortes podem também ser desafiadas ao se considerar a cognição além das modalidades dos sentidos. A questão sobre o quanto de informação específica sobre a linguagem é inata, se alguma o é, tem sido um dos tópicos dominantes na ciência cognitiva, e a literatura relevante sobre o assunto nos oferece um amplo conjunto de possibilidades. Alguns (PINKER, 1994; CHOMSKY, 1975, 1980), por exemplo, concluem que uma dotação inata específica ajuda a explicar diversos aspectos da performance linguística de crianças. Outros (COWIE, 1999; BATES, 1998; BATES *et al.*, 1995), ao invés disso, acham essa concepção pouco parcimoniosa e confusa de um ponto de vista evolucionário, e afirmam que fatores de aprendizagem não linguística condicionam e controlam de forma significativa o conjunto de possibilidades que caracterizam a linguagem falada.

A motivação para essa última concepção vem de evidências que mostram como a aquisição da linguagem depende fortemente, além da exposição ao ambiente, de uma série de fatores diferentes, incluindo a memória operacional e o desenvolvimento cognitivo geral (SEUNG; CHAPMAN, 2000). O aprendizado da primeira língua, por exemplo, é construído sobre o que as crianças já sabem sobre objetos e eventos que elas experienciaram e esse conhecimento de fundo oferece-lhes a base a partir da qual elas podem mapear palavras (*vide* E. CLARK, 2004). A habilidade delas para desenvolver uma linguagem é também afetada pela informação que elas recebem de adultos e cuidadores. Consequentemente, elas serão sensíveis e propensas a selecionar as regularidades que elas ouvem mais vezes, tais como palavras mais frequentes, sons, inflexões e construções gramaticais (SAFFRAN *et*

*al.*, 1996; DE VILLIERS, 1985). Interações sociais também parecem ser cruciais para o processo de aquisição da linguagem. Crianças aprendem mais rapidamente a nomear coisas que estão fisicamente presentes durante uma conversação e para as quais o foco de atenção conjunta é direcionado (E. CLARK, 2003; TOMASELLO, 2003). Mesmo o desenvolvimento da linguagem em crianças afetadas pelo Transtorno do Espectro Autista (TEA), geralmente considerado como tendo uma origem genética, podem ser modulados por fatores parentais e sociais, tais como gênero e alto nível de educação (GRANDGEORGE *et al.*, 2009). Isso vale para o desenvolvimento normal, no qual o tamanho e a produção do vocabulário de uma criança parecem profundamente relacionados à riqueza lexical dos seus pais, ao monitoramento das interações linguísticas e ao *status* socioeconômico (HOFF, 2003; HERR; NAIGLES, 2002).

As consequências dessas descobertas apoiam a tese do Corpo como Distribuidor e colocam ênfase principal na concepção de que propriedades do processo de aprendizagem da linguagem dependem profundamente das condições ambientais e sociais dentro das quais um indivíduo é criado. Elas também indicam que o engajamento direto com o mundo e outros indivíduos regulam o funcionamento da linguagem e que esse funcionamento é inseparavelmente conectado e explora as *affordances* da situação na qual os processos da linguagem ocorrem.

Razões adicionais para a tese do Corpo como Distribuidor surgem a partir das evidências de que diferenças nos primeiros gestos explicam disparidades no tamanho do vocabulário das crianças (ROWE; GOLDIN-MEADOW, 2009a). Pais que frequentemente usam gestos para traduzir suas palavras, oferecem às crianças uma oportunidade para aprender significados particulares pelas mãos, e esse uso de gestos de pais para filhos explica a correlação entre gestos primários das crianças aos 14 meses e as habilidades e o tamanho do vocabulário aos 54 meses. Embora o gesto não seja o único fator mediando o desenvolvimento da linguagem, evidências fortemente sugerem que a exposição a um conjunto maior de interações corporificadas determina a riqueza lexical e o crescimento do vocabulário. Outras descobertas (ROWE; GOLDIN-MEADOW, 2009b; ROWE *et al.*, 2008; OZÇALISKAN; GOLDIN-MEADOW, 2005; IVERSON; GOLDIN-MEADOW, 2005; ACREDOLO; GOODWYN, 1988) apontam na mesma direção: os pais que gesticulam com muita frequência encorajam a criança ao gesto e o gesticular pode, assim, influenciar o desenvolvimento da linguagem. Essas descobertas juntas sugerem que se a aprendizagem é baseada no corpo e correlacionada com aspectos não-linguísticos do comportamento, então apelar a algumas formas de inatismo parece inútil para explicar tais aspectos do desenvolvimento da linguagem.

Outra razão adicional para a tese do Corpo como Distribuidor vem de dados que mostram a função de gestos no raciocínio. De acordo com o enquadramento do Gesto como Ação Simulada (GSA) (HOESTETTER; ALIBABI, 2008), gestos derivam de simulações mentais de ações e de estados perceptuais que pessoas utilizam quando pensam, e eles afetam os mecanismos cognitivos a serviço de imagens mentais, juízos e solução de problemas, aumentando a ativação de áreas sensório motoras (*vide* ALIBABI *et al.*, 2014).

Deveria estar claro que nenhuma forma da Tese da Corporificação rejeita as bases biológicas do processo da linguagem e da atividade cognitiva. O que ela questiona é a adequação dos programas de pesquisa atuais que continuam a construir fortemente sobre a ideia de que o desenvolvimento linguístico e cognitivo repousa sobre os processos e mecanismos que são de domínio específico e causalmente poderosos. A ciência cognitiva corporificada produziu evidências que sugerem que estruturas não-neurais não são recursos meramente secundários. Antes, elas diversamente promovem, constituem e determinam a aquisição e o desenvolvimento de capacidades psicológicas específicas, incluindo aquelas operantes na linguagem e no processamento perceptual.

## **5. Domínios Empíricos para a Cognição Corporificada**

Nessa seção, nos concentramos sobre cinco domínios empíricos nos quais uma perspectiva corporificada motivou novos *insights* sobre a cognição e a mente: consciência visual, conceitos, memória, compreensão de outras mentes e cognição moral. Restringimos a discussão a esses cinco tópicos por razões de espaço e clareza, não porque estes sejam os únicos cinco para os quais tais ferramentas teóricas podem ser fecundamente aplicadas. Consulte Gibbs (2006) para uma apresentação extensa e discussão de uma ampla gama de aplicações.

### **5.1 Consciência Visual**

A consciência visual é normalmente concebida como um processo dentro do cérebro. Porém, as questões sobre o relacionamento entre experiência consciente e estruturas neurais são controversas empiricamente e filosoficamente. A plasticidade cerebral, por exemplo, oferece algumas razões para pensar que poderia haver



diferentes substratos neurais para uma determinada experiência visual consciente, tanto dentro do mesmo indivíduo em diferentes tempos, quanto em diferentes indivíduos ao mesmo tempo. E mesmo se alguém assume que estados do cérebro são suficientes por si mesmos para a experiência visual consciente ('a tese do substrato mínimo', em suma), pode ainda questionar se os correlatos neurais da consciência visual sistematicamente coincidem com o conteúdo da experiência visual. Embora uma quantidade considerável de neurocientistas pareça compartilhar o comprometimento com a doutrina da coincidência correlato-conteúdo, a literatura sobre os correlatos neurais da consciência, como apontam Noë e Thompson (2004), diz muito pouco ou quase nada para provar que estados neurais coincidem, no conteúdo, com a experiência visual consciente.

Assim, embora aparentemente óbvia, a concepção centrada no cérebro, endossada por eminentes estudiosos (KOCH, 2004; CHALMERS, 2000; METZINGER, 2000; CRICK; KOCH, 1990, 1998; CRICK, 1996) parece problemática após um exame mais minucioso. Defensores da cognição corporificada (NOË; THOMPSON, 2004; NOË, 2004; O'REGAN; NOË, 2001; THOMPSON; VARELA, 2001; HURLEY, 1998) colocam vários argumentos para lançar dúvidas sobre a doutrina da coincidência correlato-conteúdo e sobre a tese do substrato mínimo. Um argumento diz respeito à incomensurabilidade das características do conteúdo da experiência visual. Considerando que o conteúdo de uma experiência visual é experiencial – isto é, representado de um ponto de vista, ativo e atencional – nenhuma destas propriedades parece ser atribuível ao conteúdo de um sistema representacional neural (NOË; THOMPSON, 2004). Grosso modo, enquanto animais e pessoas experienciam o mundo relativamente a um ponto de partida egocêntrico e, fenomenologicamente falando, prestam atenção a porções dele que podem ser reveladas e exploradas através de movimentos apropriados da cabeça e do corpo, neurônios não fazem tais coisas (NOË, 2004; O'REGAN; NOË, 2001). Assim como o sentido de nossa experiência visual consciente depende do nosso domínio implícito de contingências sensorio-motoras, um conjunto de regras sobre como a estimulação sensorial varia como uma função do movimento, a experiência visual consciente é um padrão temporário de atividade habilidosa. É alguma coisa que nós fazemos (NOË; O'REGAN, 2002). Uma análise aprofundada do significado dessa afirmação, que oferece pelo menos duas interpretações principais, uma radical e uma moderada, junto com uma revisão do debate gerado, pode ser encontrada em Loughlin (2014), O'Regan (2011), Noë (2010), Shapiro (2010), Hickerson (2007). Estejam estes autores certos ou não, a afirmação é significativa, pois incita neurocientistas e filósofos a perseguir

uma abordagem bastante diferente para compreender a base da consciência visual (GANGOPADHYAY; MADARY; SPICER, 2010).

Dois notáveis fenômenos empiricamente gerados, que indicam dimensões surpreendentes para as limitações de nosso conhecimento visual explícito, foram invocados em suporte das concepções corporificadas da consciência visual. O primeiro deles, mudança à cegueira (LEVINS; SIMONS, 1997), ocorre quando as mudanças em uma cena visual são coordenadas com períodos curtos durante os quais um sujeito está mudando o foco do olhar de um ponto para outro [*saccading*]; o segundo, a cegueira desatenta (MACK; ROCK 1998; SIMONS; CHAMBRIS, 1999), quando tais mudanças ocorrem enquanto os sujeitos estão engajados em uma tarefa de atenção intensa. Sob tais condições, os sujeitos podem deixar de relatar a percepção até mesmo de mudanças massivas e marcantes (para outros observadores) em uma cena visual, tal como a presença de um gorila dançante passando pelo meio da cena.

Estes fenômenos colocam em questão a suposição tradicional de que o cérebro reconstrói modelos internos detalhados e precisos do campo visual. Essa suposição, enquanto generalizada, obscureceu dois pontos importantes, cada um dos quais motiva uma mudança para uma perspectiva corporificada da visão:

- a Visão não é um mero processo no cérebro dedicado a construir modelos mentais, mas, antes, uma habilidade do agente todo, situado e corporificado, cujos movimentos são cruciais para a agência visual (*vide* GIBSON, 1979).
- b O processo visual deveria ser reconhecido como uma atividade temporalmente estendida, onde tal atividade é guiada em parte pelo próprio agente.

O fato de que sujeitos sejam geralmente capazes de perceber amplas mudanças e inesperados elementos no campo visual revela coisas importantes. Mais interessante para os presentes propósitos, a experiência visual consciente é um engajamento habilidoso com o mundo e depende fortemente do que **fazemos** com nossos olhos, cabeça e corpo para trazer alguma coisa para dentro da consciência visual. Isso implica que devemos ‘fazer’ alguma coisa a fim de engajar em uma experiência consciente e apenas por meio de nossos movimentos corpóreos os processos de atenção podem ser dirigidos ao ambiente. Assim, corpo e mundo não

apenas importam como fontes de influência causal, mas atuam como substratos não-neurais do maquinário que realiza a enação da consciência visual.

Dado o estado atual da neurociência, a conclusão de que a experiência fenomênica não pode ser explicada por processos na cabeça parece difícil de aceitar (BLOCK, 2005). A consciência distribuída tem consequências inescapáveis. Alguém pode supor, por exemplo, que se duas pessoas com os mesmos estados internos estivessem em ambientes diferentes, suas experiências conscientes seriam diferentes, e que um cérebro numa cuba não teria qualquer experiência visual consciente, porque um cérebro numa cuba não tem corpo e, portanto, não pode interagir com o ambiente como nós normalmente interagimos.

Independentemente do quão convincentes estes argumentos são, o *insight* genuíno sobre a natureza da consciência que a ciência cognitiva corporificada gerou é que o caráter da experiência visual resulta do modo como estamos dinamicamente ligados ao mundo. Acoplamento sensorio motor com o ambiente é crucial em proporcionar ao organismo o *feedback* proprioceptivo/sinestésico necessário para o sentido de posse do movimento. Quando tocamos um objeto, por exemplo, não temos exclusivamente a experiência disso, mas enquanto tocamos e somos tocados nos experienciamos nos movendo, incluindo a sensação de controlar o próprio corpo na ação. A explicação de que a agência (o sentido de controlar o próprio corpo) se origina de processos que evoluíram para a interação com o ambiente – isto é, mecanismos para processamento sensorio e controle motor (TSAKIRIS *et al.*, 2007; BERTI *et al.*, 2005; HAGGARD, 2005; FARRER *et al.*, 2003; LEUBE *et al.*, 2003; FARRER; FRITH, 2002; CHAMINADE; DECETY, 2002) – sugere que a experiência corporificada sustenta a autoconsciência.

Uma recente abordagem contrastante da agência e de seus distúrbios relacionados, ainda que corporificada, são os trabalhos em psiquiatria fenomenológica (FUCHS, 2011, 2010, 2009, 2005; SASS; PARNAS, 2001; STANGHELLINI, 2004). Aqui, a tentativa de Frith (1992) de “neuralizar” a consciência, a subjetividade e a agência traçando-as causalmente de volta aos correlatos neurofisiológicos é desafiada por uma concepção que coloca forte ênfase sobre a noção de “corporalidade vivida”.

Consistente com a concepção de que a consciência e a ação podem ser intimamente relacionadas, estudos de imagem do cérebro mostram que ilusões de controle, frequentemente vistas em pacientes esquizofrênicos, estão associadas com uma falha no mecanismo pelo qual as consequências previstas de uma ação são conectadas às sequências pretendidas dos comandos motores (FRITH *et al.*, 2000). Deficiências desse tipo sugerem que a habilidade para controlar e ter

pensamentos conscientes pode recrutar os mesmos mecanismos empregados nas interações com o ambiente.

## 5.2 Conceitos

Uma suposição comum na abordagem tradicional é que conceitos são símbolos amodais independentes do contexto. Existem vários problemas com essa concepção e a pesquisa é enfática em sugerir que capacidades conceituais incorporam e são estruturadas em termos de padrões da atividade corpórea. Sugere-se que falar ou pensar sobre objetos implica a reativação das experiências anteriores, e o recrutamento dos mesmos circuitos neurais envolvidos durante a percepção e a ação frente a estes objetos, permitiriam a re-enação da informação multimodal (cor, tamanho, largura, etc.). A princípio, a concepção de que conceitos são representados através de símbolos abstratos, ao invés de características modais específicas, e de que a cognição exige formas estáveis de representação, deveria ser descartada ou fortemente revisada.

Evidências revelam que pessoas constroem conceitos muito diferentemente em contextos distintos (SOLOMON; BARSALOU, 2001; WISNIEWSKI, 1998; MEDIN; SHOBEN, 1988; BARSALOU; ROSS, 1986) e que a conceptualização pode variar entre indivíduos e ser diferente para o mesmo indivíduo em distintas ocasiões. Quando sujeitos são solicitados a oferecer definições para categorias, como pássaro e cadeira (*vide* BARSALOU, 1993), apenas 44% dos aspectos na definição de uma pessoa existem na definição de outra pessoa e uma grande margem de flexibilidade existe também em indivíduos que oferecem definições para as mesmas categorias duas semanas depois.

Também o padrão de interação que se tem com um objeto pode influenciar o modo em que a conceptualização é feita. Pessoas que lidam com certos itens e suas partes estruturais tendem, de modo mais frequente e extensivo do que outras pessoas, a desenvolver representações que refletem a natureza de suas interações com tais itens. Não surpreendentemente, distintos especialistas em árvores (como um taxonomista, um paisagista e um funcionário de manutenção de um parque) irão categorizá-las de maneiras que diferem uma da outra e de não-especialistas (MEDIN *et al.*, 1997). Estes estudos retratam e reforçam a teoria dos sistemas de símbolos perceptuais (BARSALOU, 1999) e fortemente indicam que mecanismos perceptuais e motores são ativados quando pessoas realizam processamentos

conceituais. Eles também sugerem que categorias completamente livres de modalidades são raras, pois conceitos, sejam concretos ou abstratos, são distribuídos sobre domínios de modalidades específicas e envolvem reativação de estados nos sistemas sensorio motores (BORONAT *et al.*, 2005; GALLESE; LAKOFF, 2005; BARSALOU *et al.*, 2003; TRANEL *et al.*, 2003; BEAUCHAMP; MARTIN, 2007; MARTIN; CHAO, 2001; MARTIN *et al.*, 2000; PULVERMÜLLER, 1999; MARTIN *et al.*, 1996).

Um suporte adicional para a teoria de que modalidades desempenham um papel fundamental na representação do conceito surge a partir do trabalho sobre verificação de propriedade (SOLOMON; BARSALOU, 2001; PECHER *et al.*, 2003). Verificação de propriedade consiste em decidir se uma propriedade em uma modalidade específica se encaixa num objeto, por exemplo, a propriedade auditiva **barulhento** para LIQUIDIFICADOR e a propriedade visual **verde** para MAÇÃ. Descobertas demonstraram que os sujeitos que realizam as tarefas respondem mais rápido e mais adequadamente quando a verificação anterior estava na mesma modalidade (por exemplo, FOLHAS – **farfalham**), ao invés de uma modalidade diferente (por exemplo, OXICOCOS [*Cranberries*] – **azedo**) (PECHER *et al.*, 2003). A eficácia é explicada ao assumir que a representação conceitual não ativa os aspectos abstratos de um objeto, mas utiliza o mesmo sistema que é recrutado pela percepção em diferentes modalidades. Assim, se o sistema auditivo é usado para ouvir o som de um liquidificador, então, para rodar uma simulação, isto é, formar um conceito, do **som** de um LIQUIDIFICADOR, o sistema auditivo será recrutado. Respostas mais lentas em diferentes modalidades são associadas com o custo da mudança de atenção, e o esforço feito na mudança de modalidades depõe contra a ideia de que o conhecimento é representado de uma maneira livre de modalidade. Fosse ao contrário, não haveria qualquer diferença entre condições de mesma modalidade e de modalidades diferentes.

Estendendo essa lógica, vários teóricos da corporificação desenvolveram a hipótese de que mesmo a representação de conceitos de emoção está baseada em estimulações corporais e exige re-enativar formas de comportamentos associadas a uma experiência original com um referente. Conceitos de emoção como “nojo”, “medo” e “raiva”, não seriam avaliações amodais e cognitivas abstratas, mas representações situadas em sensações corpóreas, incluindo expressões faciais e gestos, que geram significado (*vide* NIEDENTHAL *et al.*, 2014 para um debate recente sobre corporificação de emoções e conceitos de emoção). Uma abordagem relacionada é que a compreensão da emoção de outros baseia-se no sistema sensorio motor que fundamenta a formação do conceito de emoção em nós mesmos

(OOSTERWIJK; BARRETT, 2014).

O fato de que circuitos sensório-motores são recrutados, ou antes, **re-utilizados** para fins tais como a formação de conceito ou o processamento da linguagem, além daqueles para os quais foram estabelecidos, tal como o processamento de informação motora e sensória, favorece fortemente abordagens modais e corporificadas para a cognição em relação às abordagens amodais e abstratas. Isto também oferece uma perspectiva alternativa sobre vários tópicos nas ciências da mente, incluindo o grau de modularidade na organização cortical, o desenvolvimento do cérebro e a hipótese da localização das funções cognitivas. Confirma Anderson (2010) para uma teoria recente do re-uso neural.

Embora estas descobertas tenham provocado revisões nas abordagens tradicionais, a conclusão específica de que processos sensório motores são fisicamente constitutivos do processo conceitual não tem sido universalmente aceita com a justificativa de que os dados são consistentes com teorias diferentes (*vide* MAHON; CARAMAZZA, 2008; RUPERT, 2009b, cap. 9-10).

### 5.3 Memória

Considere o caso de lembrar como utensílios e ingredientes estão dispostos na cozinha a fim de instruir alguém a fazer um bolo. Abordagens tradicionais afirmariam que o armazenamento e a recuperação da informação deveriam ser apresentados como essencialmente independentes dos mecanismos sensório motores. Entretanto, não parece pela evidência empírica que a lembrança apela ao parentesco semântico dos ingredientes e utensílios, mas, antes, consiste em formar uma imagem que revela onde os ingredientes e os utensílios estão localizados como uma função de nosso movimento imaginado através da cozinha. A localização em si mesma serve como auxílio externo à memória e as ações corporificadas imaginadas dentro da localização proporcionam a recuperação da informação que ajuda a descobrir o que é necessário para fazer um bolo (COLE *et al.*, 1997). Que o tempo necessário para recolher e recuperar a informação é condicionado pelo esquema [*layout*] espacial imaginado com referência ao corpo do observador foi mostrado por um diferente grupo de estudos (BRYANT; WRIGHT, 1999; WALLER *et al.*, 2008; NORI *et al.* 2004; MCNAMARA 2003; WALLER *et al.*, 2002).

Os efeitos da corporificação sobre a memória também foram encontrados na realização de tarefas particulares, incluindo o raciocínio e a compreensão da

linguagem, e vários trabalhos recentes sugerem que a memória reflete diferentes capacidades corpóreas (M. WILSON, 2001; GLENBERG, 1997; CARLSON, 1997). Por exemplo, os movimentos de mãos e braços, que frequentemente acompanham a fala, não desempenham uma mera função comunicativa, mas facilitam a manutenção de representações espaciais na memória operacional (WESP *et al.*, 2001) e relembrar expressões de ação enativadas é significativamente mais fácil com respeito a codificação puramente verbal (*vide* ENGELKAMP, 1998). Esse efeito sugere que a informação motora pode ter se tornado parte do traço da memória, assim indicando que expressões de ação meramente ouvidas não produzem a mesma codificação efetiva de frases de ação realmente enativadas. A literatura empírica também suporta a hipótese de que o traço da memória inclui a postura do corpo na qual a experiência é adquirida (BARSALOU *et al.*, 2003). Com base nas condições de posturas congruente e incongruente, dados comportamentais mostram que o corpo contém a pista para lembrar eventos autobiográficos e a recuperação da memória de uma experiência passada é facilitada se a postura do corpo assumida durante esta experiência é reassumida (DIJKSTRA *et al.*, 2007). Evidência adicional sobre as capacidades de restrição do corpo foram oferecidas por Presson e Montello (1994). Nesse experimento, os sujeitos foram primeiro levados a memorizar a localização de objetos em um quarto e, então, com os olhos vendados, foram levados a apontar para os objetos. Apontar foi rápido e acurado. Enquanto alguns sujeitos foram levados a imaginar girar 90° e em seguida apontar aos objetos novamente, os outros foram levados a de fato girar 90° e em seguida apontar aos objetos. Nessa segunda metade do experimento, apontar foi lento e inacurado apenas para aqueles que imaginaram a rotação.

Uma pesquisa informativa recente sobre fenômenos empíricos relacionados à memória, que também explica em que sentido a memória pode ser compreendida como uma habilidade corporificada, a saber, um processo que incorpora a experiência corpórea, pode ser encontrado em Sutton e Williamson (2014).

## 5.4 Outras Mentes

A psicologia popular é a compreensão de senso comum que temos uns dos outros em termos de estados mentais internos, tais como crenças e desejos, que em combinação podem ser usados para prever e explicar o comportamento humano. A perspectiva tradicional sobre a compreensão da psicologia popular

pressupõe que nossas tendências atributivas aqui são geradas por uma teoria interna (PREMACK; WOODRUFF, 1978), e sob algumas versões dessa concepção “teoria-teoria” da psicologia popular, essa teoria interna é executada por um módulo de teoria da mente no cérebro (LESLIE, 1987). Oponentes dessa concepção, teóricos da simulação, minimizam a função desempenhada por um tipo de teorização abstrata típica de teorias e questionam se um tal módulo de teoria da mente existe. Eles argumentam que achados e descobertas da neurociência são consistentes com a abordagem que vê a cognição social como uma forma de simulação baseada no corpo e concebe os estados corpóreos como os blocos de construção da intersubjetividade (OBERMAN; RAMACHANDRAN, 2007; IACOBONI; DAPRETTO, 2006; RIZZOLATTI; CRAIGHERO, 2004; FERRARI *et al.*, 2003; RIZZOLATTI *et al.*, 2001; UMILTÀ *et al.*, 2001; GALLESE; GOLDMAN, 1998; RIZZOLATTI; ARBIB, 1998).

A descoberta dos neurônios espelho em células de macacos e humanos – células com propriedades sensório motoras que disparam tanto quando se realiza uma ação, quanto quando se observa a mesma ação sendo executada por outros indivíduos – foi aproveitada por teóricos da simulação e por outros oponentes da concepção teoria-teoria em suporte de suas estruturas explicativas preferidas. Para avaliações críticas dos estudos fMRI e PET, visando mostrar a presença de um sistema de espelhamento neuronal em humanos, confira Turella *et al.* (2009) e, também, Hickok (2008).

Evidência indireta em suporte de um sistema de espelhamento neuronal em humanos vem de estudos sobre a reatividade do ritmo cerebral, o ritmo alfa posterior e o ritmo mu central, durante a observação da ação. Enquanto o ritmo mu está presente durante o descanso motor, ele desaparece quando movimentos ativos são realizados. Resultados eletrofisiológicos mostraram que observar a ação executada por outro indivíduo bloqueia o ritmo mu do observador, assim oferecendo evidência para um sistema ressonante, que conecta a ação observada com a ação do repertório motor do próprio sujeito (OBERMAN *et al.*, 2005; COCHIN *et al.*, 1998; GASTAUT; BERT, 1954; COHEN-SEAT *et al.*, 1954). Outro corpo de evidência em suporte da existência de um sistema combinatório observação/execução em humanos vem dos estudos de estimulação magnética transcraniana (FADIGA *et al.*, 1995). O córtex motor esquerdo de participantes normais foi estimulado tanto durante a observação de movimentos de braços intransitivos quanto de movimentos dirigidos a objetivos. Potenciais motores evocados foram registrados a partir de vários músculos das mãos. Os resultados mostraram um aumento seletivo nos potenciais motores evocados em regiões que participantes normalmente recrutam para produzir



a ação observada. Isto significa que quando observamos uma pessoa pegar um copo de café, a mesma população neural que controla a execução dos movimentos de pegar se torna ativa nas áreas motoras do observador.

Segundo essa concepção, é a imitação corporificada do corpo observado na ação que diretamente nos permite reconhecer outras pessoas como nós, e não um processo abstrato, inferencial e algo como teórico. A hipótese de que a compreensão da ação é baseada em um mecanismo de ressonância não exclui a possibilidade de que outros processos, baseados em descrições de movimento, poderiam influenciar essa função. Isto enfatiza simplesmente a primazia de uma combinação direta, automática e pré-reflexiva entre a observação e a execução da ação. Ao aceitar essa premissa a tensão tradicional entre agir e pensar encolheu consideravelmente, pois a capacidade para detectar o significado do comportamento de outros consiste em empregar os mesmos recursos usados para moldar nosso comportamento motor.

Em uma contribuição recente, Craighero (2014) revisa as principais descobertas da literatura sobre neurônios espelho e discute as funções possíveis de um sistema dedicado a ações de codificação, incluindo a possibilidade que ele oferece para a aquisição de novas habilidades motoras.

## 5.5 Cognição Moral

Uma tradição dominante na teoria moral nos últimos séculos coloca a razão no centro do pensamento e do comportamento moral. O trabalho de Kohlberg (1969), de desenvolvimento cognitivo sobre cognição moral, reflete esse espírito. Kohlberg consistentemente endossou, ao longo das linhas do cognitivismo clássico, um modelo racionalista no qual emoções e estados do corpo podem ser levados em consideração pela razão, mas é o raciocínio puro que no fim das contas leva a decisões morais. Seu paradigma inspirou a maioria dos estudos líderes na área, todos caracterizados pela concepção comum de que o processamento cognitivo no domínio moral é livre da economia das emoções e do corpo.

Trabalhos recentes na cognição moral corporificada desafiam esse paradigma. Modelos do intuicionismo social, por exemplo, afirmam que muitos juízos morais parecem ser o resultado de componentes afetivos (GREENE; HAIDT, 2002; HAIDT *et al.*, 1993). Sujeitos apresentados a descrições de ações sem danos, mas propensas a produzir uma forte resposta/reação afetiva (por exemplo, comer o

cachorro de estimação depois que um carro o matou) geralmente julgam que a ação descrita pela frase está errada. Quando pressionados para oferecer uma justificção, tais sujeitos normalmente se concentram em danos não existentes associados às ações, indicando que o raciocínio consciente não é um bom preditor de seus juízos sobre estar errado. Se reações afetivas desempenham uma função pervasiva em juízos morais, uma função que escapa da (e é de fato mascarada pela) reflexão consciente, isso sugere que a cognição moral é moldada e condicionada por “sentimentos viscerais”, ao invés de ser o produto do raciocínio abstrato.

Exemplos adicionais que apoiam a natureza corporificada da cognição moral vêm da literatura experimental que trata especificamente do nojo/repugnância (LERNER *et al.*, 2004; WHEATLEY; HAIDT, 2005). O nojo envolve componentes físicos fortes e mudanças corpóreas explícitas, tais como náusea, estômago revirado, garganta apertada, vontade de vomitar, expelir com a língua e franzir o nariz. Mesmo que o nojo tenha evoluído como uma reação relacionada à comida, indicando que uma substância deveria ser ou evitada ou expelida, ele é também uma emoção de rejeição social, estendendo seu alcance e influência no domínio da cognição moral (NIEDENTHAL *et al.*, 2005; PRINZ, 2004).

Experimentos recentes mostram, por exemplo, que a sensação de repugnância induzida pela exposição a um mau cheiro ou a sujeira vista num quarto, torna os juízos morais mais fortes (SCHNALL *et al.*, 2008a), e que os indivíduos que se limpam fisicamente consideram que certas ações morais são menos erradas do que os participantes não expostos a uma manipulação de limpeza (SCHNALL *et al.*, 2008b). Estes efeitos de uma “aspereza moral” podem ser induzidos até mesmo para avaliações morais sobre certas questões ou grupos. Em um estudo recente (INBAR *et al.*, 2009), pesquisadores descobriram que pessoas fazem avaliações mais negativas sobre gays e lésbicas quando expostas a um odor nocivo no quarto do que quando o odor não está presente. Até a raiva mostrou moldar profundamente as representações disponíveis ao juízo moral. A raiva sobre um incidente no trânsito antes de chegar ao trabalho pode aumentar a confiança em preconceitos ao entrevistar um candidato numa posterior entrevista de emprego (DESTENO *et al.*, 2004).

Crescentes evidências de que percepção, emoção e juízo são baseados em mecanismos sensório motores motivou a tese da especificidade do corpo (CASASANTO, 2011, 2009; DE LA VEJA *et al.*, 2012; BRUNYÉ *et al.*, 2012), a afirmação de que pessoas com diferentes tipos de corpos tendem a pensar diferentemente. Destros, por exemplo, conceitualizam “bom” e “mau” em termos de

dominância motora ao invés das convenções sociais e tendem a preferir ou fazer avaliações positivas de pessoas e de objetos apresentados ao seu lado direito (CASASANTO, 2014, 2009; CASASANTO; HENETZ, 2012). Esse mapeamento implícito da valência do espaço pode ser manipulado via mudanças na interação corpo-ambiente. Indivíduos que perderam o uso de sua mão dominante ou que vestem uma luva que atrapalha a fluência motora mostram uma inversão em suas associações de valência do espaço (CASASANTO; CHRYSIKOU, 2011). Para um debate detalhado de como juízos afetivos incorporam padrões de interação corpórea com o ambiente e são condicionados por um conjunto de capacidades corporificadas, confira Casasanto (2014).

A literatura empírica sugere que a especificidade das pistas corpóreas e reações afetivas próprias de alguém (por exemplo, náusea, excitação) guiam e condicionam o processo cognitivo em domínios sociais e morais. Além disso, a literatura sugere que graves déficits acontecem quando indivíduos não exibem e não fazem uso dessas pistas e reações. A hipótese do marcador somático de Damasio (1994, 1996) afirma que estados corpóreos, normalmente acionados durante as experiências emocionais, são re-enativados sempre que certas situações acontecem ou são consideradas, e tal re-enação exclui consequências deletérias do curso da ação de alguém. Quando a capacidade para integrar estas sensações (sejam positivas ou negativas) com o próprio conhecimento dos fatos é gravemente comprometida, como é o caso de pacientes com lesão no córtex pré-frontal ventromedial (VMPFC), a realização de juízos e decisões é seriamente debilitada. Como uma 'tarefa de jogo' (BECHARA *et al.*, 1994) revela, na ausência de estados corporificados (por exemplo, resposta antecipatória de condução da pele), pacientes VMPFC perdem uma fonte fundamental de informação sobre possíveis resultados associados a ações diferentes (*vide* BECHARA *et al.*, 2000). Se a hipótese de Damasio está correta, então o pano de fundo afetivo e corpóreo implicado em vários tipos de juízo moral não simplesmente leva a compreensões diferentes e conceptualizações sobre a situação em questão, mas é parte do maquinário físico que realiza os processos cognitivos.

Enquanto pensamos que há amplo suporte empírico para a ideia de que a atividade sensorial motora e o processamento cognitivo central são mais profundamente dependentes entre si do que anteriormente pensávamos, e que a concepção de que a atividade corpórea pode restringir, distribuir e regular a atividade neural, a ciência cognitiva corporificada tem também sido invocada em apoio a ideias filosóficas mais radicais sobre a cognição e a mente. Por exemplo, como

mencionamos na Seção 3, tem se argumentado que a cognição corporificada implica que a cognição em si mesma escapa para dentro do corpo, e, finalmente para o ambiente (A. CLARK, 2008). Defensores da concepção tradicional de que a cognição é limitada pelo crânio argumentam, como réplica, que essa inferência é enganosa e que a visão sugerida é implausível e metafisicamente extravagante (RUPERT, 2009b; ADAMS; AIZAWA, 2008; AIZAWA, 2007). Abordamos essas questões filosóficas adicionais na Seção 6 abaixo.

## 6. Divisões mais nítidas na Cognição Corporificada

A diferença que a cognição corporificada faz nas três questões discutidas na Seção 4 – a modularidade da mente, a natureza da representação mental e o nativismo sobre a mente – permanece uma questão viva de debate em filosofia da mente e em ciência cognitiva. Isso vale para a interpretação de resultados empíricos particulares descritos na Seção 5.

Pensamos que algumas dessas discordâncias refletem e contribuem para divisões mais nítidas sobre a significação da ciência cognitiva corporificada. Vamos discutir quatro dessas questões nessa seção final, estruturando nossa discussão em torno de quatro perguntas correspondentes:

- 1 Quais ganhos para a pesquisa empírica a ciência cognitiva corporificada oferece?
- 2 Em que medida os achados gerados pela ciência cognitiva corporificada podem ser acomodados pelas ferramentas da ciência cognitiva tradicional?
- 3 Qual é a relação entre cognição corporificada e a tese da mente estendida?
- 4 Que implicações tem para a agência, o *self* e a subjetividade, aceitar a natureza corporificada da cognição?

Nosso propósito aqui não é oferecer respostas extensas para essas questões, mas indicar brevemente o que nossa revisão dos trabalhos contemporâneos sobre cognição corporificada sugere sobre as questões que eles levantam. Tomamos uma de cada vez.

## 6.1 Ganhos para a pesquisa empírica

Dado que a ciência cognitiva corporificada tem suas origens em uma variedade de insatisfações com a ciência cognitiva tradicional, ela explorou novas questões sobre a cognição e gerou resultados inesperados em alguns casos. Como vimos na seção anterior, a ciência cognitiva corporificada continua a produzir pesquisa empírica que é interessante, nova e controversa. Nesse ponto, a ciência cognitiva corporificada não é simplesmente (ou principalmente) um mantra filosófico vazio de conteúdo empírico, mas um grupo de perspectivas sobre a cognição cuja orientação e enraizamento empírico não pode ser questionado.

Mas há uma posição alternativa que se pode tomar sobre essa questão que é mais discreta sobre o ganho empírico da “cognição corporificada”. Embora não haja dúvida sobre a “energia” [*oomph*] da ciência cognitiva corporificada, o modo pelo qual esse trabalho desafia concepções tradicionais e exige certas formas determinadas da Tese da Corporificação que articulamos – Corpo como Restrição, Corpo como Distribuidor e Corpo como Regulador – poderia ser questionado. Por exemplo, Lawrence Shapiro (2011) argumentou que as concepções de Lakoff e Johnson sobre metáfora, pensamento e corpo são inteiramente compatíveis com os princípios centrais da ciência cognitiva tradicional, tal como a ideia de que a cognição centralmente envolve computação sobre representações mentais internas (*vide* SHAPIRO, 2010). Robert Rupert (2009b) defende de modo mais geral a compatibilidade entre os achados empíricos da ciência cognitiva corporificada e as hipóteses centrais da ciência cognitiva tradicional. Do mesmo modo, Fred Adams (2010) argumentou que se deveria distinguir entre a premissa empírica nos argumentos para a corporificação da cognição e a premissa lógico-metafísica exigida, e que esta última é raramente sustentada. Adams se concentra no trabalho de Glenberg sobre o significado de *affordances* (GLENBERG; KASCHAK, 2002; GLENBERG; ROBERTSON, 2000, GLENBERG *et al.*, 2005), mas suas afirmações são bastante gerais. Dessa perspectiva, deve-se separar o trigo empírico da ciência cognitiva corporificada do joio filosófico revolucionário que caracteriza este movimento desde o início. Essa questão, por sua vez, nos leva à próxima questão.

## 6.2 Acomodação pela ciência cognitiva tradicional

Em geral, questionar se a ciência cognitiva corporificada tem resultados na frente revolucionária não passa por esboçar considerações gerais – digamos, da subdeterminação da teoria pelos dados – mas, por uma consideração detalhada de resultados empíricos particulares (*vide* RUPERT, 2009b, cap.11). Por sua natureza, esse tipo de argumento, cuja necessidade endossamos, consome tempo e espaço, especialmente dada a diversidade de trabalhos que caem sob o rótulo “cognição corporificada”. Isso é maior na avaliação desse tipo de desafio aos proponentes da cognição corporificada. Aqui nos contentamos em fazer um ponto geral sobre o estado da dialética presente, e apontar onde acreditamos que o ônus da prova reside.

Suponha que simplesmente admitimos a afirmação histórica de que o foco e a orientação da ciência cognitiva tradicional não tomam a cognição como sendo dependente, de qualquer forma significativa, do corpo. O que isso implica sobre o poder explicativo da ciência cognitiva tradicional frente a Tese da Corporificação? Lembre-se que analisamos a Tese da Corporificação em termos de três teses determinadas sobre a natureza da dependência da cognição sobre o corpo, cada uma com suas próprias consequências particulares:

**Corpo como Restrição:** o corpo de um agente funciona de modo a restringir significativamente a natureza e o conteúdo das representações processadas pelo sistema cognitivo desse agente.

**Corpo como Distribuidor:** o corpo de um agente funciona de modo a distribuir carga computacional e representacional entre estruturas neurais e não-neurais.

**Corpo como Regulador:** o corpo de um agente funciona de modo a regular a atividade cognitiva sobre espaço e tempo, garantindo que cognição e ação sejam intimamente coordenadas.

Aqueles que buscam resistir ao desafio que uma ou mais dessas concepções colocam à ciência cognitiva tradicional têm duas opções fundamentais: negar a verdade da tese correspondente ou rejeitar a(s) inferência(s) feitas a partir dessa tese, acerca do que a ciência cognitiva tradicional pode e não pode explicar.

Defensores da ciência cognitiva tradicional tem considerável poder dialético disponível aqui e eles têm feito uso efetivo de uma estratégia argumentativa familiar

para resistir ao desafio da cognição corporificada: admitir que há um sentido fraco e limitado no qual uma ou mais dessas teses particulares da corporificação é correta, mas argumentar que a lacuna inferencial entre tais teses e a rejeição de concepções, tais como o computacionalismo e o representacionalismo, não é transponível (ADAMS, 2010; RUPERT, 2009a, 2009b), uma estratégia que aqueles criticamente simpáticos à ciência cognitiva corporificada tem também feito uso eficaz (SHAPIRO, 2010, 2011). Nessa medida, o ônus da prova de fato recai diretamente sobre os proponentes da ciência cognitiva corporificada que sustentam que a promessa revolucionária da cognição corporificada é efetiva em mostrar como aquelas lacunas podem ser transpostas.

### 6.3 Cognição corporificada e a tese da mente estendida

Uma das implicações supostamente radicais da cognição corporificada é a Tese da Mente Estendida, que diz que a mente de um agente e os processos associados não são nem limitados pelo crânio e nem mesmo limitados pelo corpo, mas se estendem ao mundo do agente. Diferente da Tese da Corporificação, essa tese surge mais explicitamente das discussões filosoficamente orientadas – do funcionalismo (HARMAN, 1988), do computacionalismo e individualismo (R. WILSON, 1994, 1995, cap. 3-4) e da crença (CLARK; CHALMERS, 1998). Por isso, ela tem um tipo diferente de relacionamento com o trabalho empírico na ciência cognitiva em relação à ideia da cognição corporificada. Apesar dos legados destas histórias diferentes, como notamos na Seção 1, a cognição corporificada e a mente estendida recentemente passaram a ser vistas como ervilhas numa mesma vagem, como variantes da ciência cognitiva situada.

O primeiro ponto para chamar a atenção é que nada em qualquer uma das três formas determinadas da Tese da Corporificação implica a tese da mente estendida. Assim, a concepção de que a cognição é corporificada (em algum sentido específico: restrita, distribuída e regulada) é compatível com a rejeição da concepção de que a cognição é estendida. Expresso em termos da tese do Corpo como Distribuidor, talvez o processamento cognitivo seja distribuído pelo corpo através de substratos neurais e não-neurais, mas todas as fontes não-neurais relevantes **estão contidas dentro dos limites do corpo**. Acreditamos que essa é a posição ocupada pelo núcleo da comunidade da ciência cognitiva corporificada. Por exemplo, a análise recente no artigo de Barsalou [2008], sobre “cognição enraizada” [*grounded*

*cognition*], omite completamente a menção a qualquer grande literatura filosófica sobre cognição estendida, é indicativa desse estado de coisas.

Segundo, muitas das mais influentes defesas da tese da mente estendida apelam a considerações apenas tangencialmente relacionadas ao corpo – ao computacionalismo e individualismo (R. WILSON, 1994), à cognição distribuída e em nível de grupo (HUTCHINS, 1995), ao princípio de paridade (CLARK; CHALMERS, 1998) e à realização (R. WILSON, 2001, 2004, cap. 5-6). Por essa razão, o debate sobre estes argumentos e a concepção estendida da mente que eles supostamente apoiam foram apenas recentemente agrupados por defensores (R. WILSON, 2010; A. CLARK, 2003, 2008) e por críticos (ADAMS; AIZAWA, 2008; RUPERT, 2009b). Essa recente consideração conjunta é mutuamente benéfica para ambas as discussões sobre cognição corporificada e estendida.

Desse modo, e terceiro, apesar de sua independência, afirma-se que os argumentos mais poderosos para uma dessas concepções também oferecem fortes razões para aceitar uma outra. Por exemplo, Andy Clark (2008) argumenta a partir da ativa corporificação da cognição para a tese da mente estendida. Similarmente, algumas das objeções mais incisivas a uma dessas concepções também parecem servir como base para rejeitar a outra. Por exemplo, críticos da tese da mente estendida, tais como Adams e Aizawa (2008) e Rupert (2004, 2009b), objetam que aqueles que defendem a tese confundem ou omitem a distinção entre **causas** externas da cognição e **constituintes** externos da cognição. Essa acusação de cometer uma “falácia do acoplamento-constituição” é também prontamente feita contra concepções particulares da cognição corporificada, tal como a concepção de Alva Nöe (2004) de que a experiência perceptual é constituída por habilidades sensório motoras (*vide* PRINZ, 2009; AIZAWA, 2007; RUPERT, 2006; BLOCK, 2005). Ainda que possa haver diferenças relevantes entre a cognição corporificada e a estendida que sugerem que tais argumentos e objeções não se transferem, existem ao menos padrões, considerações de paridade que colocam o ônus da prova sobre aqueles que reivindicam essas diferenças.

Quarto, pode haver razões mais profundas para pensar que a cognição corporificada e a cognição estendida ficam em pé ou caem juntas. Rupert (2009b), por exemplo, recentemente argumentou contra ambas, a cognição corporificada e a estendida, em parte criando um caso positivo para o que ele chama de concepção de *sistemas cognitivos* sobre os limites da cognição, e que essa concepção sugere, junto com nossa melhor ciência empírica, que a cognição começa e termina no cérebro. Se Rupert está correto, então a cognição não é nem corporificada e nem



estendida porque ambas as concepções são incompatíveis com uma abordagem independentemente motivada da natureza cerebral das arquiteturas cognitivas integradas.

Inversamente, ao apresentar uma concepção geral da cognição situada como extensão cognitiva, Wilson e Clark (2009, p. 56) afirmam que “muitas formas de cognição corporificada, propriamente compreendida, acabam por envolver justamente os tipos de extensão cognitiva que articulamos aqui”, uma nota promissória que um de nós (R. WILSON, 2010) tentou pagar ao oferecer o seguinte argumento explícito ligando cognição corporificada e estendida:

- 1 A função de alguns processos visuais é guiar a ação através da informação visual.
- 2 Um modo fundamental para realizar esta função é através da corporificação ativa do processo visual (*à la* Corpo como Distribuidor).
- 3 Processos visuais são ativamente corporificados nesse mesmo sentido se em sua operação normal em ambientes naturais, estes processos são acoplados com atividades corpóreas na medida em que formam um sistema integrado com ganho funcional.
- 4 Mas processos visuais que são corporificados ativamente, nesse sentido, são também estendidos. Assim,
- 5 Alguns processos visuais, e os sistemas visuais que estes processos constituem fisicamente, são estendidos.

Claramente, como a premissa que explicitamente traça uma conexão entre as teses da Corporificação e da Mente Estendida, (4) é onde esse argumento será examinado por aqueles céticos sobre a conclusão do argumento. Se isso pode ser feito enquanto se aceita (1)-(3), entretanto, é obscuro e o tipo de questão exige mais atenção nesse debate.

#### **6.4 Agência, *self* e subjetividade**

Se a mente não está limitada ao crânio, mas é ao menos corporificada e talvez ainda estendida, então qual concepção deveríamos adotar sobre o *self*, a

subjetividade e a consciência? No penúltimo parágrafo, Clark e Chalmers (1998) defendem a concepção de que, dizendo coloquialmente, aonde a mente vai, o *self* vai atrás: se a mente é estendida, por exemplo, o *self* também o é (ver também A. CLARK, 2001, 2003). Já que muito do que importa para a identidade do próprio *self* é cognitivo por natureza, ao menos nas concepções tradicionais sobre o *self* e a identidade pessoal, essa concepção “*self* segue a mente” parece um padrão natural.

Se os limites do *self* mudam com os da mente, saindo do crânio em direção ao corpo, ou mesmo do corpo ao mundo, como a concepção *self*-segue-a-mente sugere, então aceitar a cognição corporificada ou estendida terá consequências interessantes sobre autonomia, sociabilidade, identidade pessoal e responsabilidade. Por exemplo, pode ser que em alguns casos interferir no espaço peripessoal [*peripersonal space*] de alguém, o espaço perto do corpo, ou ainda em certos pertences de alguém, terá um significado moral comparável a interferir no corpo de uma pessoa. E como Clark e Chalmers (1998) indicaram em seu parágrafo final, certas formas de atividade social podem ser reconcebidas como um tipo de atividade do pensamento. A distribuição social da tomada de decisão humana também mitigaria a responsabilidade individual por uma transgressão, produzindo assim ramificações dramáticas para nossas práticas de atribuição de culpabilidade legal. De fato, se as forças situacionais e as contingências ambientais desempenham uma função fisicamente constitutiva no processo de tomada de decisão, de modo a se tornar realizadores parciais de seu próprio comportamento, como a assim chamada “Hipótese do Controle Frágil” parece sugerir (para críticas a essa concepção, consulte Churchland e Suler [2009]), então, seres humanos teriam pouco ou nenhum controle sobre suas ações e provavelmente nenhuma competência normativa.

Contra a concepção *self*-segue-a-mente, Wilson (2004, p. 141-143) argumentou que mesmo se alguém aceita que a mente é estendida, existem razões para resistir à ideia de que o *self* é igualmente estendido. Essa resistência retorna precisamente aos tipos de implicações indicadas acima e suas consequências frequentemente radicais, profundamente contraintuitivas e intrigantes. Por exemplo, se sujeitos de cognição (agentes ou indivíduos) são em si mesmos distribuídos entre cérebro, corpo e mundo, então por que deveríamos punir somente uma parte desse indivíduo, a parte dentro do corpo, quando ele comete um crime? (Considere isso uma **redução** truncada.) Mais amplamente, ainda que agentes como sujeitos da cognição não sejam apenas um conjunto de circuitos neurais e experiências corpóreas, reformular a agência e a subjetividade dentro do quadro de referências estendido provavelmente exige uma reconceitualização muito mais abrangente e

um tanto incômoda de noções como competência normativa, liberdade, controle e identidade pessoal. Talvez isso simplesmente nos diga que muito mais exploração é necessária sobre como experiências corporificadas nos contextos do mundo real moldam o processamento cognitivo. Ou talvez isto sugere que estratégias mais conservadoras deveriam ser empregues para abordar o que sujeitos de processamentos cognitivos realmente são.

Uma tal estratégia é apelar para a distinção trivial entre **o sujeito ou agente da cognição**, que pode ser prontamente identificado como estando onde o local do controle para um dado sistema cognitivo está, alojado no corpo do agente, e os *sistemas cognitivos* nos quais o processo cognitivo é realizado, os quais são frequentemente estendidos (R. WILSON, 2004, cap. 5-6). Uma tal distinção foi antes usada para dar sentido a *sistemas biológicos* estendidos, tal como aranhas e as teias que elas tecem – estes organismos são limitados, grosseiramente falando, por seus corpos orgânicos coesos, mas ainda atuam no mundo através dos sistemas biológicos estendidos que eles constroem (R. WILSON, 2005, cap. 1-4). Assim, um apelo a essa distinção aqui não é *ad hoc* e oferece um princípio de base para uma concepção mais conservadora e tradicional da agência e do *self* dentro de um quadro de referência da cognição estendida.

## Referência bibliográfica

- ACREDOLO, L.; GOODWYN, S. Symbolic gesturing in normal infants. **Child Development**, v. 59, p. 450-466, 1988.
- ADAMS, F. **The Bounds of Cognition**. Malden: Blackwell Publishing, 2008.
- ADAMS, F. Embodied Cognition. **Phenomenology and Cognition**, v. 9, n. 4, p. 619-628, 2010.
- ADAMS, F.; AIZAWA, K. Why the Mind is Still in the Head. *In*: ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 78-95.
- AGRE, P.; CHAPMAN, D. Pengi: An implementation of a theory of activity. *In*: **The Proceedings of the Sixth National Conference on Artificial Intelligence**. Seattle, WA: American Association for Artificial Intelligence, 1987, p. 268-272.
- AIZAWA, K. Understanding the Embodiment of Perception. **Journal of Philosophy**, v. 104, p. 5-25, 2007.

- ALIBALI, M.; BONCODDO, R.; HOSTETTER, A. Gesture in Reasoning. In: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014.
- ANDERSON, M. Embodied Cognition: A Field Guide. **Artificial Intelligence**, v. 149, n. 1, p. 91-130, 2003.
- ANDERSON, M. On the Grounds of (X)-Grounded Cognition. In: CALVO, P.; GOMILA, T. (ed.). **The Elsevier Handbook of Cognitive Science: An Embodied Approach**, p. 423-435.
- ANDERSON, M. Neural re-use as a fundamental organizational principle of the brain. **Behavioral and Brain Sciences**, v. 33, n. 4, p. 245-313, 2010.
- BAILLARGEON, R. The Acquisition of Physical Knowledge in Infancy: A Summary in Eight Lessons. In: GOSWAMI, U. (Ed.). **Blackwell Handbook of Childhood Cognitive Development**. Oxford: Blackwell, 2002. p. 47-83.
- BAILLARGEON, R.; SPELKE, E.; WASSERMAN, S. Object Permanence in Five-Month-Old Infants. **Cognition**, v. 20, n. 3, p. 191-208, 1985.
- BALLARD, D. Animate vision. **Artificial Intelligence Journal**, v. 48, p. 57-86, 1991.
- BARON-COHEN, S. **Mindblindness: An Essay on Autism and Theory of Mind**. Cambridge, MA: MIT Press, 1995.
- BARSALOU, L. Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In: COLLINS, A.; GATHERCOLE, S.; CONWAY, M. (ed.). **Theories of memory**. Hillsdale, NJ: Erlbaum, 1993. p. 29-101.
- BARSALOU, L. Perceptual Symbol Systems. **Behavioral and Brain Sciences**, v. 22, p. 577-609, 1999.
- BARSALOU, L. Abstraction in perceptual symbol systems. **Philosophical Transactions of the Royal Society of London: Biological Sciences**, v. 358, p. 1177-1187, 2003.
- BARSALOU, L. Grounded Cognition. **Annual Review of Psychology**, v. 59, p. 617-645, 2008.
- BARSALOU, L. Simulation, situated conceptualization, and prediction. **Philosophical Transactions of the Royal Society of London: Biological Sciences**, v. 364, p. 1281-1289, 2009.
- BARSALOU, L.; ROSS, B. The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. **Journal of Experimental Psychology: Learning, Memory, and Cognition**, v. 12, p. 116-134, 1986.

- BARSALOU, L. *et al.* Grounding conceptual knowledge in modality-specific systems. **Trends in Cognitive Sciences**, v. 7, p. 84-91, 2003.
- BATES, E. Nativism, empiricism, and the origins of knowledge. **Infant Behaviour and Development**, v. 21, n. 2, p. 181-200, 1998.
- BATES, E. Plasticity, Localization and Language Development. *In*: BROMAN, S. H.; FLETCHER, J. M. (ed.). **The Changing Nervous System: Neurobehavioural Consequences of Early Brain Disorders**. New York: Oxford University Press, 2001. p. 214-253
- BATES, E.; DALE, P.; THAL, D. Individual Differences and their implications for theories of language development. *In*: FLETCHER, P.; MACWHINNEY, B. (ed.). **Handbook of Child Language**. Oxford: Blackwell, 1995. p. 96-151.
- BEAUCHAMP, M.; MARTIN, A. Grounding object concepts in perception and action: Evidence from fMRI studies of tools. **Cortex**, v. 43, p. 461-468, 2007.
- BECHARA, A. *et al.* Insensitivity to future consequences following damage to human prefrontal cortex. **Cognition**, v. 50, p. 7-15, 1994.
- BECHARA, A.; DAMASIO, H.; DAMASIO, A. Emotion, Decision Making and the Orbitofrontal Cortex. **Cerebral Cortex**, v. 10, p. 295-307, 2000.
- BECHTEL, W. Explanation, Mechanism, Modularity, and Situated Cognition. *In*: ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 155-170.
- BEER, R. **Intelligence as Adaptive Behavior**. New York: Academic Press, 1990.
- BEER, R. Dynamical approaches to cognitive science. **Trends in Cognitive Sciences**, v. 4, p. 91-99, 2000.
- BEER, R. The Dynamics of active categorical perception in an evolved model agent. **Adaptive Behavior**, v. 11, p. 209-243, 2003.
- BERTI, A. *et al.* Shared Cortical Anatomy for Motor Awareness and Motor Control. **Science**, v. 309, p. 488-491, 2005.
- BLACK, M. **Models and Metaphors**. Ithaca: Cornell University Press, 1962.
- BLOCK, N. Review of Alva Noë Action in Perception. **Journal of Philosophy**, v. 102, p. 259-272, 2005.
- BORGHI, A.; GLENBERG, A.; KASCHAK, M. Putting words in perspective. **Memory & Cognition**, v. 32, n. 6, p. 863-873, 2004.
- BORONAT, C. *et al.* Distinction between manipulation and function knowledge of objects: Evidence from functional magnetic resonance imaging. **Cognitive Brain Research**, v. 23, p. 361-373, 2005.

- BROOKS, R. Intelligence without representation. **Artificial Intelligence**, v. 47, p. 139-159, 1991a.
- BROOKS, R. Intelligence without reason. *In: Proceedings of 12th International Joint Conference on Artificial Intelligence*. Sydney, Australia: 1991b, p. 569-595.
- BROOKS, R. **Flesh and Machine: How Robots Will Change Us**. New York: Pantheon, 2002.
- BRUNYÉ, T. et al. Body-specific representations of spatial location. **Cognition**, v. 123, n. 2, p. 229-239, 2012.
- BRYANT, D.; WRIGHT, G. How body asymmetries determine accessibility in spatial function. **Quarterly Journal of Experimental Psychology**, v. 52A, p. 487-508, 1999.
- CARLSON, R. **Experienced Cognition**. Mahwah, NJ: Erlbaum, 1997.
- CARRUTHERS, P. **The architecture of the mind**. Oxford: Clarendon Press, 2006.
- CASASANTO, D. Embodiment of abstract concepts: good and bad in right- and left-handers. **Journal of Experimental Psychology (General)**, v. 138, n. 3, p. 351-367, 2009.
- CASASANTO, D. Different Bodies, Different Minds The Body Specificity of Language and Thought. **Current Directions in Psychological Science**, v. 20, n. 6, p. 378-383, 2011.
- CASASANTO, D. Bodily relativity. *In: SHAPIRO, L. (ed.). The Routledge Handbook of Embodied Cognition*. London: Routledge, 2014.
- CASASANTO, D.; CHRYSIKOU, E. When Left Is "Right" Motor Fluency Shapes Abstract Concepts. **Psychological Science**, v. 22, n. 4, p. 419-422, 2011.
- CASASANTO, D.; HENETZ, T. Handedness shapes children's abstract concepts. **Cognitive Science**, v. 36, n. 2, p. 359-372, 2012.
- CHALMERS, D. What is a neural correlate of consciousness?. *In: METZINGER, T. (ed.). Neural Correlates of Consciousness: Conceptual and Empirical Questions*. Cambridge, MA: MIT Press, 2000.
- CHAMINADE, T.; DECETY, J. Leader or follower? Involvement of the inferior parietal lobule in agency. **NeuroReport**, v. 13, p. 1975-1978, 2002.
- CHAO, L.; MARTIN, A. Representation of manipulable man-made objects in the dorsal stream. **NeuroImage**, v. 12, p. 478-484, 2000.
- CHEMERO, T. **Radical Embodied Cognitive Science**. Cambridge, MA: MIT Press, 2009.
- CHOMSKY, N. **Reflections on Language**. New York: Pantheon, 1975.
- CHOMSKY, N. **Rules and Representations**. New York: Columbia University Press, 1980.

- CLARK, A. **Being There: Putting Mind, Body, and World Together Again**. Cambridge, MA: MIT Press, 1997.
- CLARK, A. Reasons, Robots, and the Extended Mind. **Mind and Language**, v. 16, p. 121-145, 2001.
- CLARK, A. **Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence**. New York: Oxford University Press, 2003a.
- CLARK, A. **Supersizing the Mind: Embodiment, Action, and Cognitive Extension**. New York: Oxford University Press, 2008.
- CLARK, A.; CHALMERS, D. The Extended Mind. **Analysis**, v. 58, p. 10-23, 1998.
- CLARK, A.; TORIBIO, J. Doing without representing? **Synthese**, v. 101, p. 401-431, 1994.
- CLARK, E. **First Language Acquisition**. Cambridge: Cambridge University Press, 2003b.
- CLARK, E. How language acquisition builds on cognitive development. **Trends in Cognitive Science**, v. 8, n. 10, p. 472-478, 2004.
- COCHIN, S. et al. Perception of motion and qEEG activity in human adults. **Electroencephalography Clinical Neurophysiology**, v. 107, p. 287-295, 1998.
- COHEN-SEAT, G. et al. Etudes expérimentales de l'activité nerveuse pendant la projection cinématographique. **Review International de Filmologie**, v. 5, p. 7-64, 1954.
- COLE, M.; HOOD, L.; MCDERMOTT, R. Concepts of ecological validity: Their differing implications for comparative cognition. *In*: COLE, M.; ENGESTROEM, Y. (ed.). **Mind, culture, and activity**. New York: Cambridge University Press, 1997. p. 48-58.
- COSMIDES, L.; TOOBY, J. The Modular Nature of Human Intelligence. *In*: SCHEIBEL, A.; SCHOPF, J. (ed.). **The Origin and Evolution of Intelligence**. Sudbury, MA: Jones; Bartlett Publishers, 1997. p. 71-101.
- COWIE, F. **What's Within: Nativism Reconsidered**. New York: Oxford University Press, 1999.
- CRAIGHERO, L. The Role of the Motor System in Cognitive Functions. *In*: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014.
- CRAIN, S.; PIETROSKI, P. Nature, nurture and universal grammar. **Linguistic and Philosophy**, v. 24, p. 139-186, 2001.
- CRICK, F. Towards a Neurobiological Theory of Consciousness. **Seminars in Neurosciences**, v. 2, p. 263-275, 1990.
- CRICK, F. Visual perception: Rivalry and consciousness. **Nature**, v. 379, p. 485-486, 1996.

- CRICK, F.; KOCH, C. Consciousness and neuroscience. **Cerebral Cortex**, v. 8, p. 97-107, 1998.
- DAMASIO, A. **Descartes' Error: Emotion, Reason, and the Human Brain**. New York: Putnam Publishing, 1994.
- DAMASIO, A. The somatic marker hypothesis and the possible functions of the prefrontal cortex. **Philosophical Transaction: Biological Sciences**, v. 351, p. 1413-1420, 1996.
- DEGENAAR, J.; O'REGAN, J.; FORTHCOMING. Sensorimotor Theory and Enactivism. **Topoi**, v. 36, p. 393-407, 2017. DOI: <https://doi.org/10.1007/s11245-015-9338-z>
- DE LA VEGA, I. *et al.* Emotional valence and physical space: limits of interaction. **Journal of Experimental Psychology (Human Perception and Performance)**, v. 38, n. 2, p. 375-385, 2012.
- DESTENO, D. *et al.* Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. **Psychological Science**, v. 15, p. 319-324, 2004.
- DE VILLIERS, J. Learning how to use verbs: lexical coding and the influence of input. **Journal of Child Language**, v. 12, p. 587-595, 1985.
- DIJKSTRA, K.; KASCHAK, M.; ZWAAN, R. Body posture facilitates retrieval of autobiographical memories. **Cognition**, v. 102, p. 139-149, 2007.
- DONALD, M. **Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition**. Cambridge, MA: Harvard University Press, 1991.
- DREYFUS, H. **What Computers Can't Do**. New York: Harper & Row, 1972.
- ELMAN, J. *et al.* **Rethinking Innateness: A Connectionist Perspective on Development**. Cambridge, MA: MIT Press, 1996.
- ENGELKAMP, J. **Memory for actions**. Hove, England: Psychology Press, 1998.
- FADIGA, L. *et al.* Motor facilitation during action observation: a magnetic stimulation study. **Journal of Neurophysiology**, v. 73, p. 2608-2611, 1995.
- FARRER, C. *et al.* Modulating the experience of agency: a positron emission tomography study. **NeuroImage**, v. 18, p. 324-333, 2003.
- FARRER, C.; FRITH, C. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. **NeuroImage**, v. 15, p. 596-603, 2002.
- FERRARI, P. *et al.* Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. **European Journal of Neuroscience**, v. 17, p. 1703-1714, 2003.
- FODOR, J. **The Language of Thought**. Cambridge, MA: Harvard University Press, 1975.



- FODOR, J. Methodological Solipsism Considered as a Research Strategy in Cognitive Science. **Behavioral and Brain Sciences**, v. 3, p. 63-73, 1980.
- FODOR, J. **Representations**. Cambridge, MA: MIT Press, 1981.
- FODOR, J. **The Modularity of Mind**. Cambridge, MA: MIT Press, 1983.
- FODOR, J. **Psychosemantics**. Cambridge, MA: MIT Press, 1987.
- FODOR, J. **Concepts: Where Cognitive Science Went Wrong**. Cambridge, MA: MIT Press, 1997.
- FODOR, J. **The Mind Doesn't Work That Way**. Cambridge, MA: MIT Press, 2000.
- FODOR, J. A.; PYLYSHYN, Z. W. Connectionism and cognitive architecture: A critical analysis. **Cognition**, v. 28, p. 3-71, 1988.
- FRITH, C. **The cognitive neuropsychology of schizophrenia**. Hillsdale: Earlbaum, 1992.
- FRITH, C.; BLAKEMORE, S.; WOLPERT, D. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. **Brain Research Reviews**, v. 31, p. 257-363, 2000.
- FUCHS, T. Corporealized and Disembodied Minds: A Phenomenological View of the Body in Melancholia and Schizophrenia. **Philosophy, Psychiatry, and Psychology**, v. 12, n. 2, p. 95-107, 2005.
- FUCHS, T. Embodied Cognitive Neuroscience and its Consequences for Psychiatry. **Poiesis and Praxis**, v. 6, n. 3-4, p. 219-233, 2009.
- FUCHS, T. The Psychopathology of Hyperreflexivity. **Journal of Speculative Philosophy**, v. 24, n. 3, p. 239-255, 2010.
- FUCHS, T. Are Mental Illnesses Diseases of the Brain? *In*: PH.D, S.; PH.D, J. (ed.). **Critical Neuroscience**. Chichester: Wiley-Blackwell, 2011. p. 331-344.
- GALLAGHER, S. **How the Body Shapes the Mind**. Oxford: Oxford University Press, 2005.
- GALLAGHER, S. Philosophical Antecedents of Situated Cognition. *In*: ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. [s.l.] Cambridge University Press, 2009. p. 35-51.
- GALLAGHER, S.; ZAHAVI, D. **The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science**. New York: Routledge, 2008.
- GALLESE, V. *et al.* Action recognition in the premotor cortex. **Brain**, v. 119, p. 593-609, 1996.
- GALLESE, V.; GOLDMAN, A. Mirror neurons and the simulation theory of mind-reading. **Trends in Cognitive Sciences**, v. 12, p. 493-501, 1998.
- GALLESE, V. *et al.* Perception through action. **Psyche**, v. 5, p. 5-21, 1999.
- GALLESE, V.; LAKOFF, G. The brain's concepts: The role of the sensorimotor system in conceptual knowledge. **Cognitive Neuropsychology**, v. 21, p. 455-479, 2005.

- GANGOPADHYAY, N.; MADARY, M.; SPICER, F. (ed.). **Perception, Action and Consciousness**. New York: Oxford University Press, 2010.
- GASTAUT, H.; BERT, J. EEG changes during cinematographic presentation. **Electroencephalography Clinical Neurophysiology**, v. 6, p. 433-444, 1954.
- GAUTHIER, I. *et al.* Expertise for cars and birds recruits brain areas involved in face recognition. **Nature Neuroscience**, v. 3, p. 191-197, 2000.
- GIBBS, R. **Embodiment and Cognitive Science**. Cambridge: Cambridge University Press, 2006.
- GIBSON, J. **The Ecological Approach to Visual Perception**. Boston: Houghton Mifflin, 1979.
- GLENBERG, A. What memory is for. **Behavioral and Brain Science**, v. 20, p. 1-55, 1997.
- GLENBERG, A. Indexical understanding of instructions. **Discourse Processes**, v. 28, p. 1-26, 1999.
- GLENBERG, A. *et al.* Episodic affordances contribute to language comprehension. **Language and Cognition**, v. 1, n. 1, p. 113-135, 2009.
- GLENBERG, A. *et al.* Grounding Language in Bodily States: The Case for Emotion. *In*: THINKING, R.; PECHER, D. (ed.). **The Grounding of Cognition: The Role of Perception and Action in Memory, Language**. Cambridge: Cambridge University Press, 2005. p. 115-128.
- GLENBERG, A.; KASCHAK, M. Grounding language in action. **Psychonomic Bulletin and Review**, v. 9, p. 558-565, 2002.
- GLENBERG, A.; ROBERTSON, D. Symbol grounding and meaning, A comparison of high- dimensional and embodied theories of meaning. **Journal of Memory and Language**, v. 43, p. 379-401, 2000.
- GRANDGEORGE, M. *et al.* Environmental factors influence language development in children with autism spectrum disorders. **PLoS ONE**, v. 4, n. 4, p. 4683, 2009.
- GREENE, J.; HAIDT, J. How (and where) does moral judgment work? **Trends in Cognitive Sciences**, v. 6, n. 12, p. 517-523, 2002.
- GREGORY, R. **Eye and Brain**. London: Weidenfeld; Nicolson, 1966.
- GRIFFITHS, P.; STOTZ, K. How the mind grows: a developmental perspective on the biology of cognition. **Synthese**, v. 122, p. 29-51, 2000.
- HAGGARD, P. Conscious intention and motor cognition. **Trends in Cognitive Science**, v. 9, n. 6, p. 290-295, 2005.
- HAIDT, J.; KOLLER, S.; DIAS, M. Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog. **Journal of Personality and Social Psychology**, v. 65, n. 4, p. 613-628, 1993.

- HAITH, M. Who put the cog in infant cognition: Is rich interpretation too costly? **Infant Behavior and Development**, v. 21, p. 167-179, 1998.
- HARMAN, G. Wide Functionalism. *In*: SCHIFFER, S.; STEELE, D. (ed.). **Cognition and Representation**. Boulder, CO: Westview Press, 1988.
- HARNAD, S. The Symbol Grounding Problem. **Physica D**, v. 42, n. 4, p. 335-346, 1990.
- HICKERSON, R. Perception as Knowing How to Act: Alva Noë's Action in Perception. **Philosophical Psychology**, v. 20, n. 4, p. 505-517, 2007.
- HICKOK, G. Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. **Journal of Cognitive Neuroscience**, v. 21, n. 7, p. 1229-1243, 2008.
- HIRSCHFELD, L.; GELMAN, S. *In*: EDS. (ed.). **Mapping the mind: Domain specificity in cognition and culture**. New York: Cambridge University Press, 1994.
- HOFF, E. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. **Child Development**, v. 74, n. 5, p. 1368-1378, 2003.
- HOFF, E.; NAIGLES, L. How children use input to acquire a lexicon. **Child Development**, v. 73, n. 2, p. 418-433, 2002.
- HOSTETTER, A.; ALIBALI, M. Visible embodiment: Gestures as simulated action. **Psychonomic Bulletin & Review**, v. 15, n. 3, p. 495-514, 2008.
- HURLEY, S. **Consciousness in Action**. London: Harvard University Press, 1998.
- HURLEY, S.; NOË, A. Neural Plasticity and Consciousness. **Biology and Philosophy**, v. 18, p. 131-168, 2003.
- HUSSERL, E. **Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy**. First Book: General Introduction to a Pure Phenomenology. Tradução: F. Kersten. The Hague: Nijhoff, 1913.
- HUSSERL, E. **Cartesian Meditations**. Tradução: D. Cairns. Dordrecht: Kluwer, 1931.
- HUTCHINS, E. **Cognition in the Wild**. Cambridge, MA: MIT Press, 1995.
- HUTTO, D.; MYIN, E. **Radicalizing Enactivism: Basic Minds Without Content**. Cambridge, MA: MIT Press, 2013.
- IACOBONI, M.; DAPRETTO, M. The mirror neuron system and the consequences of its dysfunction. **Nature Review Neuroscience**, v. 7, p. 942-951, 2006.
- INBAR, Y.; PIZARRO, D.; BLOOM, P. Disgusting smells cause decreased liking of homosexuals. **Emotion**, v. 12, n. 1, p. 23-7, 2012. DOI: 10.1037/a0023984
- IVERSON, J.; GOLDIN-MEADOW, S. Gesture paves the way for language development. **Psychological Science**, v. 16, p. 368-371, 2005.

- JOHNSON, M. **The Body in Mind**. Chicago: University of Chicago Press, 1987.
- JOHNSON, M. Functional Brain Development in Infants: Elements of an Interactive Specialization Framework. **Child Development**, v. 71, n. 1, p. 75-81, 2000.
- JOHNSON, M. **The Meaning of the Body**. Chicago: University of Chicago Press, 2007.
- KANWISHER, N. Domain specificity in face perception. **Nature Neuroscience**, v. 3, p. 759-763, 2000.
- KANWISHER, N.; MCDERMOTT, J.; CHUN, M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. **The Journal of Neuroscience**, v. 17, p. 4302-4311, 1997.
- KARMILOFF-SMITH, A. Precis of Beyond modularity: A developmental perspective on cognitive science. **Behavioral and Brain Sciences**, v. 17, n. 4, p. 693-745, 1994.
- KASCHAK, M.; GLENBERG, A. Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. **Journal of Memory and Language**, v. 43, p. 508-529, 2000.
- KLÖPPEL, S. et al. Can Left-Handedness be Switched? Insights from an Early Switch of Handwriting. **The Journal of Neuroscience**, v. 27, n. 29, p. 7847-7853, 2007.
- KOCH, C. **The Quest for Consciousness: A Neurobiological Approach**. Denver, CO: Roberts & Company Publishers, 2004.
- KOHLBERG, L. Stage and sequence: The cognitive-developmental approach to socialization. *In*: GOSLIN, D. (ed.). **Handbook of socialization theory and research**. Chicago: Rand McNally, 1969. p. 347-480.
- LAKOFF, G. **Women, Fire, and Dangerous Things: What Categories Reveal About the Mind**. Chicago: University of Chicago Press, 1987.
- LAKOFF, G.; JOHNSON, M. **Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought**. New York: Basic Books, 1999.
- LAKOFF, G.; JOHNSON, M. **Metaphors We Live By**. Chicago: University of Chicago Press, 1980.
- LEEUWEN, L.; SMITSMAN, A.; LEEUWEN, C. Affordances, perceptual complexity, and the development of tool use. **Journal of Experimental Psychology: Human Perception and Performance**, v. 20, p. 174-191, 1994.
- LERNER, J.; SMALL, D.; LOEWENSTEIN, G. Heart strings and purse strings: Carryover effects of emotions on economic decisions. **Psychological Science**, v. 15, p. 337-341, 2004.

- LESLIE, A. Pretence and representation: The origins of "theory of mind". **Psychological Review**, v. 94, p. 412-426, 1987.
- LEUBE, D. et al. The neural correlates of perceiving one's own movements. **NeuroImage**, v. 20, p. 2084-2090, 2003.
- LEVIN, D.; SIMONS, D. Failure to detect changes to attended objects in motion pictures. **Psychonomic Bulletin and Review**, v. 4, p. 501-506, 1997.
- LI, Y. et al. Experience with moving visual stimuli driver the early development of cortical direction selectivity. **Nature**, v. 456, p. 952-956, 2008.
- LOUGHLIN, V. Sensorimotor Knowledge and the Radical Alternative. In: BISHOP, J.; MARTIN, A. (ed.). **Contemporary Sensorimotor Theory**. Dordrecht: Springer International Publishing, 2014. p. 105-116.
- MACHERY, E. Concept empiricism: A methodological critique. **Cognition**, v. 104, p. 19-46, 2007.
- MACIVER, M. Neuroethology: From Morphological Computation to Planning. In: ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 480-504.
- MACK, A.; ROCK, I. **Inattentional Blindness**. Cambridge, MA: MIT Press, 1998.
- MAHON, B.; CARAMAZZA, A. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual knowledge. **Journal of Physiology**, v. 102, p. 59-70, 2008.
- MARR, D. **Vision: A Computational View**. San Francisco: Freeman Press, 1982.
- MARTIN, A.; CHAO, L. Semantic memory and the brain: structure and process. **Current Opinion in Neurobiology**, v. 11, p. 194-201, 2001.
- MARTIN, A.; UNGERLEIDER, L.; HAXBY, J. Category-specificity and the brain: the sensory-motor model of semantic representations of objects. In: GAZZANIGA, M. (ed.). **Category Specificity and the Brain: The Sensory-Motor Model of Semantic Representations of Objects**. Cambridge, MA: MIT Press, 2000. p. 1023-1036.
- MARTIN, A. et al. Neural correlates of category specific knowledge. **Nature**, v. 379, p. 649-652, 1996.
- MCNAMARA, T. How are the locations of objects in the environment represented in memory? In: FRESKA, C.; BRAUER, W.; HABEL, C.; WENDER, K. F. (ed.). **Spatial Cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning**. Berlin: Springer-Verlag, 2003. p. 174-191

- MCNEILL, D. **Hand and Mind: What Gestures Reveal about Thought**. Chicago: University of Chicago Press, 1992.
- MEDIN, D.; LYNCH, E.; COLEY, J. Categorization and Reasoning among tree Experts: Do All Roads Lead to Rome? **Cognitive Psychology**, v. 32, p. 49-96, 1997.
- MEDIN, D.; SHOBEN, E. Context and structure in conceptual combination. **Cognitive Psychology**, v. 20, p. 158-190, 1988.
- MENARY, R. (ed.). **The Extended Mind**. Cambridge, MA: MIT Press, 2010.
- MERLEAU-PONTY, M. **Phenomenology of Perception**. Tradução: C. Smith. London: Routledge; Kegan Paul, 1945.
- METZINGER, T. (ed.). **Neural Correlates of Consciousness**. Empirical and Conceptual Questions. Cambridge, MA: MIT Press, 2000.
- MYIN, E.; DEGENAAR, J. Enactive Vision. In: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014.
- NEWELL, A.; SIMON, H. **Human problem solving**. Englewood Cliffs, NJ: Prentice-Hall, 1972.
- NIEDENTHAL, P. *et al.* Embodiment in Attitudes, Social Perception, and Emotion. **Personality and Social Psychology Review**, v. 9, p. 184-211, 2005.
- NIEDENTHAL, P.; WOOD, A.; RYCHLOWSKA, M. Embodied Emotion Concepts. In: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014. p. 240-249.
- NOË, A. **Action in Perception**. Cambridge, MA: MIT Press, 2004.
- NOË, A. Vision without representation. In: GANGOPADHYAY, N.; MADARY, M.; SPICER, F. (ed.). **Perception, Action, and Consciousness: Sensorimotor Dynamics and Two Visual Systems**. New York: Oxford University Press, 2010.
- NOË, A.; O'REGAN, K. On the Brain-basis of Visual Consciousness: A Sensorimotor Account". In: NOË, A.; THOMPSON, E. (ed.). **Vision and Mind: Selected Readings in the Philosophy of Perception**. Cambridge, MA: MIT Press, 2002.
- NÖE, A.; THOMPSON, E. Are there neural correlates of consciousness? **Journal of Consciousness Studies**, v. 11, n. 1, p. 2-28, 2004.
- NORI, R.; INCHINI, T.; GIUSBERTI, F. Object localization and frames of reference. **Cognitive Processing**, v. 5, n. 1, p. 45-53, 2004.
- OBERMAN, L. *et al.* EEG evidence for mirror neuron dysfunction in autism spectrum disorders. **Cognitive Brain Research**, v. 24, p. 190-198, 2005.
- OBERMAN, L.; RAMACHANDRAN, V. The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative

- deficits of autism spectrum disorders. **Psychological Bulletin**, v. 133, n. 2, p. 310-327, 2007.
- OOSTERWIJK, S.; BARRETT, L. Embodiment in the Construction of Emotion Experience and Emotion Understanding. *In*: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014.
- O'REGAN, J. **Why Red Doesn't Sound Like a Bell**: Understanding the feel of consciousness. Oxford: Oxford University Press, 2011.
- O'REGAN, J.; NOË, A. A sensorimotor account of vision and visual consciousness. **Behavioral and Brain Sciences**, v. 25, n. 4, p. 883-975, 2001.
- ORTONY, A. (ed.). **Metaphor and Thought**. Cambridge: Cambridge University Press, 1979.
- OUDEJAMS, R. *et al.* The relevance of action in perceiving affordances: Perception and the catchableness of fly balls. **Journal of Experimental Psychology: Human Perception and Performance**, v. 22, p. 879-891, 1996.
- OZÇALIŞKAN, S.; GOLDIN-MEADOW, S. Do parents lead their children by the hand? **Journal of Child Language**, v. 32, n. 3, p. 481-505, 2005.
- PAOLO, D.; EZEQUIEL; THOMPSON, E. The enactive approach. *In*: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014.
- PASCUAL-LEONE, A.; HAMILTON, R. The metamodal organization of the brain. **Progress in Brain Research**, v. 134, p. 427-445, 2001.
- PASCUAL-LEONE, A. *et al.* Study and modulation of human cortical excitability with transcranial magnetic stimulation. **Journal of Clinical Neurophysiology**, v. 15, n. 4, p. 333-343, 1998.
- PECHER, D. Verifying conceptual properties in different modalities produces switching costs. **Psychological Science**, v. 14, p. 119-124, 2003.
- PECHER, D.; ZEELBERG, R.; BARSALOU, L. Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior of prior activation. **Psychonomic Bulletin & Review**, v. 11, n. 1, p. 164-167, 2004.
- PECHER, D.; ZWAAN, R. (ed.). **Grounding cognition**. The role of perception and action in memory, language, and thinking. Cambridge: Cambridge University Press, 2005.
- PINKER, S. **The Language Instinct**: How the Mind Creates Language. New York: Harper Collins, 1994.
- PORT, R.; GELDER, T. **Mind as Motion**. Cambridge, MA: MIT Press, 1995.

- PREMACK, D.; WOODRUFF, G. Does the chimpanzee have a theory of mind? **Behavioural and Brain Science**, v. 4, p. 515-526, 1978.
- PRESSON, C.; MONTELLO, D. Updating after rotational and translational body movements: Coordinate structure of perspective space. **Perception**, v. 23, p. 1447-1455, 1994.
- PRINZ, J. **Gut reactions: A perceptual theory of emotion**. New York: Oxford University Press, 2004.
- PRINZ, J. Is Consciousness Embodied? *In*: ROBBINS, P.; AYDELE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 419-436
- PULVERMÜLLER, F. Words in the brain's language. **Behavioral and Brain Sciences**, v. 22, p. 253-336, 1999.
- QUARTZ, S.; SEJNOWSKI, T. The neural basis of cognitive development: A constructivist manifesto. **Behavioural and Brain Sciences**, v. 20, p. 537-596, 1997.
- RIZZOLATTI, G.; ARBIB, M. Language within our grasp. **Trends in Neuroscience**, v. 21, p. 188-194, 1998.
- RIZZOLATTI, G.; CRAIGHERO, L. The Mirror-Neuron System. **Annual Review of Neuroscience**, v. 27, p. 169-192, 2004.
- RIZZOLATTI, G. *et al.* Premotor cortex and the recognition of motor actions. **Cognitive Brain Research**, v. 3, p. 131-41, 1996.
- RIZZOLATTI, G.; FOGASSI, L.; GALLESE, V. Neurophysiological mechanisms underlying the understanding and imitation of action. **Nature Neuroscience Review**, v. 2, p. 661-670, 2001.
- ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. New York: Cambridge University Press, 2010.
- ROCK, I. **The logic of perception**. Cambridge, MA: MIT Press, 1983.
- ROCK, I. **Indirect Perception**. Cambridge, MA: MIT Press, 1997.
- ROWE, M. Early gesture selectively predicts later language learning. **Developmental Science**, v. 12, n. 1, p. 182-187, 2009.
- ROWE, M.; GOLDIN-MEADOW, S. Differences in early gesture explain SES disparities in child vocabulary size at school entry. **Science**, v. 323, n. 5916, p. 951-953, 2009.
- ROWE, M.; ÖZÇALISKAN, S.; GOLDIN-MEADOW, S. Learning words by hand: Gesture's role in predicting vocabulary development. **First Language**, v. 28, n. 2, p. 182-199, 2008.



- ROWLANDS, M. **The Body in Mind**. New York: Cambridge University Press, 1999.
- ROWLANDS, M. Situated Representation. *In*: ROBBINS, P.; AYDENE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 117-133
- RUPERT, R. Challenges to the Hypothesis of Extended Cognition. **Journal of Philosophy**, v. 101, n. 8, p. 389-428, 2004.
- RUPERT, R. Review of Raymond W Gibbs, Jr., Embodiment and Cognitive Science. **Notre Dame Philosophical Reviews**, v. 8, 2006.
- RUPERT, R. Innateness and the Situated Mind. *In*: ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009a. p. 96-116.
- RUPERT, R. **Cognitive Systems and the Extended Mind**. Oxford: Oxford University Press, 2009b.
- SADATO, N. *et al.* Neural networks for Braille reading by the blind. **Brain**, v. 121, p. 1213-1229, 1998.
- SADATO, N. *et al.* Activation of the primary visual cortex by Braille reading in blind subjects. **Nature**, v. 380, p. 526-528, 1996.
- SAFFRAN, J.; ASLIN, R.; NEWPORT, E. Statistical learning by 8-month-old infants. **Science**, v. 274, p. 1926-1928, 1996.
- SARTRE, J. **Being and Nothingness**. Tradução: H.E. Barnes. New York: Philosophical Library, 1943.
- SASS, L.; PARNAS, J. Phenomenology of Self-Disturbances in Schizophrenia: Some Research Findings and Directions. **Philosophy, Psychiatry, and Psychology**, v. 8, n. 4, p. 347-356, 2001.
- SCHNALL, S.; BENTOS, J.; HARVEY, S. With a clean conscience. Cleanliness reduces the severity of moral judgments. **Psychological Science**, v. 19, n. 12, p. 1219-1222, 2008.
- SCHNALL, S. *et al.* Disgust as Embodied Moral Judgment. **Personality and Social Psychology Bulletin**, v. 34, n. 8, p. 1096-1109, 2008.
- SEUNG, H.; CHAPMAN, R. Digit span in individuals with Down syndrome and typically developing children: Temporal aspects. **Journal of Speech, Language, and Hearing Research**, v. 43, p. 609-620, 2000.
- SHAPIRO, L. A Clearer Vision. **Philosophy of Science**, v. 64, p. 131-153, 1997.
- SHAPIRO, L. Embodied Cognition. *In*: MARGOLIS, E.; SAMUELS, R.; STICH, S. (ed.). **Oxford Handbook of Philosophy and Cognitive Science**. Oxford: Oxford University Press, 2010.

- SHAPIRO, L. **Embodied Cognition**. New York: Routledge, 2011.
- SIMON, H. **Machine as mind**. Cambridge, MA: MIT Press, 1995. p. 23-40
- SIMONS, D.; CHABRIS, C. Gorillas and our midst: Sustained inattentive blindness for dynamic events. **Perception**, v. 28, p. 1059-1074, 1999.
- SMITH, B. Situatedness/Embeddedness. *In*: WILSON, R. A.; KEIL, F. C. (ed.). **The MIT Encyclopedia of the Cognitive Sciences**. Cambridge, MA: MIT Press, 1999. p. 769-770.
- SMITH, L.; THELEN, E. Development as dynamic system. **Trends in Cognitive Science**, v. 7, n. 8, p. 343-348, 2003.
- SOLOMON, K.; BARSALOU, L. Representing properties locally. **Cognitive Psychology**, v. 43, p. 129-169, 2001.
- SPELKE, E. *et al.* Origins of knowledge. **Psychological Review**, v. 99, p. 605-632, 1992.
- SPELKE, E.; VISHTON, P.; HOFSTEN, C. Object perception, object-directed action, and physical knowledge in infancy. *In*: GAZZANIGA, M. (ed.). **The Cognitive Neurosciences**. Cambridge, MA: MIT Press, 1995.
- SPERBER, D. In Defense of massive modularity. *In*: DUPOUX, E. (ed.). **Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler**. Cambridge, MA: MIT Press, 2001. p. 47-57.
- SPIVEY, M. *et al.* Eye movements during comprehension of spoken scene descriptions. **Proceedings of the 22nd Annual Conference of the Cognitive Science Society Meeting**, 2000, p. 487-492.
- STANGHELLINI, G. **Disembodied Spirits and Deanimated Bodies: The Psychopathology of Common Sense**. New York: Oxford University Press, 2004.
- SUCHMAN, L. **Plans and Situated Action**. Cambridge: Cambridge University Press, 1987.
- SUHLER, C.; CHURCHLAND, P. Control: conscious and otherwise. **Trends in Cognitive Science**, v. 13, n. 8, p. 341-347, 2009.
- SUTTON, J.; WILLIAMSON, K. Embodied Remembering. *In*: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. London: Routledge, 2014.
- THELEN, E.; L.B., S. **A dynamic systems approach to the development of cognition and action**. Cambridge, MA: MIT Press, 1994.
- THOMPSON, E. **Colour Vision: A Study in Cognitive Science and the Philosophy of Perception**. New York: Routledge, 1995.
- THOMPSON, E. **Mind and Life**. Cambridge, MA: Harvard University Press, 2007a.
- THOMPSON, E. **Mind in Life: Biology, Phenomenology, and the Sciences of Mind**. Cambridge, MA: Harvard University Press, 2007b.

- THOMPSON, E.; PALACIOS, A.; VARELA, F. Ways of coloring: Comparative color vision as case study for cognitive science. **Behavioral and Brain Sciences**, v. 15, p. 1-25, 1992.
- THOMPSON, E.; VARELA, F. Radical Embodiment: Neural Dynamics and Consciousness. **Trends in Cognitive Sciences**, v. 5, p. 418-425, 2001.
- TOMASELLO, M. **Constructing a Language**. Harvard University Press, 2003.
- TRANEL, D. *et al.* Neural correlates of conceptual knowledge for actions. **Cognitive Neuropsychology**, v. 20, p. 409-432, 2003.
- TSAKIRIS, M. *et al.* Neural signatures of body ownership: A sensory network for bodily self-consciousness. **Cerebral Cortex**, v. 17, p. 2235-2244, 2007.
- TURELLA, L. *et al.* Mirror neurons in humans: Consistent or confounding evidence? **Brain & Language**, v. 108, p. 10-21, 2009.
- UMILTÀ, M. *et al.* I know what you are doing: A neurophysiological study. **Neuron**, v. 32, p. 91-101, 2001.
- VAN GELDER, T. What might cognition be if not computation? **Cognitive Science Research Report**, v. 75, Indiana University, 1992.
- VARELA, F.; THOMPSON, E.; ROSCH, E. **The Embodied Mind: Cognitive Science and Human Experience**. Cambridge, MA: MIT Press, 1991.
- WAGMAN, G.; CARELLO, C. Affordances and inertial constraints in tool use. **Ecological Psychology**, v. 13, p. 173-195, 2001.
- WALLER, D.; LIPPA, Y.; RICHARDSON, A. Isolating observer-based reference directions in human spatial memory: Head, body, and the self-to-array axis. **Cognition**, v. 106, p. 157-183, 2008.
- WALLER, D. *et al.* Orientation Specificity and Spatial Updating of Memories for Layouts. **Journal of Experimental Psychology: Learning, Memory, and Cognition**, v. 28, n. 6, p. 1051-1063, 2002.
- WARREN, W. Perceiving affordances: Visual guidance of stair climbing. **Journal of Experimental Psychology: Human Perception and Performance**, v. 10, p. 683-703, 1894.
- WESP, R. *et al.* Gestures maintain spatial imagery. **American Journal of Psychology**, v. 114, p. 591-600, 2001.
- WHEATLEY, T.; HAIDT, J. Hypnotically induced disgust makes moral judgments more severe. **Psychological Science**, v. 16, p. 780-784, 2005.
- WHEELER, M. **Reconstructing the Cognitive World**. Cambridge, MA: MIT Press, 2005.
- WILSON, M. The case for sensorimotor coding in working memory. **Psychonomic Bulletin and Review**, v. 9, p. 49-57, 2001a.

- WILSON, M. Six views of embodied cognition. **Psychonomic Bulletin and Review**, v. 9, p. 625-636, 2002.
- WILSON, R. Wide Computationalism. **Mind**, v. 103, p. 351-372, 1994.
- WILSON, R. **Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind**, Cambridge Studies in Philosophy. Cambridge: Cambridge University Press, 1995.
- WILSON, R. Two Views of Realization. **Philosophical Studies**, v. 104, p. 1-30, 2001b.
- WILSON, R. **Boundaries of the Mind**. The Individual in the Fragile Sciences: Cognition. Cambridge: Cambridge University Press, 2004.
- WILSON, R. What Computers (Still, Still) Can't Do: Jerry Fodor on Computation and Modularity. *In*: STANTON, R. J.; EZCURDIA, M.; VIGER, C.D. (ed.). **New Essays in Philosophy of Language and Mind**, p. 407-425, 2005.
- WILSON, R. The Drink You Have When You're Not Having a Drink. **Mind and Language**, v. 23, p. 273-283, 2008.
- WILSON, R. **Extended Vision**. New York: Oxford University Press, 2010.
- WILSON, R.; CLARK, A. How to Situate Cognition: Letting Nature Take its Course. *In*: AYDEDE, M.; ROBBINS, P. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 55-77.
- WILSON, R.; KEIL, F. (ed.). **The MIT Encyclopedia of the Cognitive Sciences**. Cambridge, MA: MIT Press, 1999.
- WINOGRAD, T.; FLORES, F. **Understanding Computers and Cognition**. Norwood, NJ: Understanding Computers; Cognition, 1986.
- WISNIEWSKI, E. Property instantiation in conceptual combination. **Memory & Cognition**, v. 26, p. 1330-1347, 1998.
- WU, L.; BARSALOU, L. Perceptual simulation in conceptual combination: Evidence from property generation. **Acta Psychologica**, v. 132, p. 173-189, 2009.
- YANG, S.; GALLO, D.; BEILOCK, S. Embodied Memory Judgments: A Case of Motor Fluency. **Journal of Experimental Psychology: Learning, Memory, and Cognition**, v. 35, n. 5, p. 1359-1365, 2009.
- ZWAAN, R.; YAXLEY, R. Hemispheric difference in semantic-relatedness judgments. **Cognition**, v. 87, n. 3, p. 79-86, 2003.

## (VI) Emoção\*

Autores: Andrea Scarantino e Ronald de Sousa

Tradução: Thiago Gruner (PPGFIL-UFRGS)

Revisão: Eros Moreira de Carvalho (UFRGS)

Nenhum aspecto de nossa vida mental é mais importante para a qualidade e o sentido de nossa existência do que as emoções. Elas são aquilo que tornam a vida digna de ser vivida e, por vezes, digna de terminar. Assim, não surpreende que a maioria dos grandes filósofos clássicos possuía célebres teorias das emoções. Tais teorias tipicamente concebem as emoções como respostas fenomenologicamente salientes de um sujeito a eventos importantes, e como capazes de dispararem certos comportamentos e mudanças corporais. Mas o que, **sim**, surpreende é que ao longo de boa parte do século XX, cientistas e filósofos da mente tenham no mais das vezes negligenciado as emoções – em parte por causa da alegria comportamentalista

---

\*SCARANTINO, A.; DE SOUZA, R. Emotion. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2018. Disponível em: <https://plato.stanford.edu/archives/win2018/entries/emotion/>. Acesso em: 30 out. 2021.

The following is the translation of the entry on Emotion by Andrea Scarantino and Ronald de Sousa, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/win2018/entries/emotion/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/emotion/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.

a estados mentais internos e em parte porque a variedade de fenômenos cobertos pela palavra "emoção" desencoraja uma teorização metódica. Em décadas recentes, contudo, as emoções novamente se tornaram foco de um vigoroso interesse na filosofia e na ciências afetivas. Nosso objetivo nesse verbete é dar conta desses desenvolvimentos, concentrando-nos sobretudo na questão descritiva do que são as emoções, mas também lidando com a questão normativa de se as emoções são racionais. À luz das intensas trocas entre pesquisadores de diferentes campos, deixou de ser vantajoso falarmos de filosofia da emoção em separado de outras abordagens disciplinares, particularmente da psicologia, neurociência e biologia evolucionária. É por isso que fizemos o esforço de prestar suficiente atenção aos avanços científicos, convencidos de que a fecundidade transdisciplinar é nossa melhor chance de fazer progressos na teoria da emoção.

Após algumas observações metodológicas breves com a intenção de esclarecer o que diferencia a abordagem filosófica da perspectiva mais geral das ciências cognitivas, começamos por apresentar algumas das maneiras encontradas pelos pesquisadores para conceber o lugar da emoção na topografia da mente. Notaremos que as emoções têm sido historicamente concebidas de três maneiras: como experiências, como avaliações, e como motivações. Cada uma dessas tradições de pesquisa captura algo de verdadeiro e significativo sobre as emoções, mas nenhuma teoria dentro de qualquer tradição parece ser imune a contraexemplos e casos problemáticos. Quanto à racionalidade das emoções, distinguiremos duas variedades dela – a racionalidade cognitiva e a racionalidade estratégica – e exploraremos uma variedade de maneiras segundo as quais as emoções falham ou têm êxito em relação a diferentes padrões de racionalidade.

## 1. Definindo as emoções: quais são os *desiderata*?

Dois grandes *desiderata* têm governado o projeto de definir as emoções tanto na filosofia quanto nas ciências afetivas: (a) obter a compatibilidade com o uso linguístico comum; e (b) obter uma fecundidade teórica. Uma definição que visa exclusivamente (a) é uma definição **descritiva**. Uma definição que visa (b), às custas de talvez violar algumas intuições comuns, é **prescritiva**. A fim de garantir a compatibilidade com a linguagem comum, filósofos tradicionais têm se apoiado em introspecção, experimentos de pensamento, observações ocasionais, *insights* colhidos de textos literários e outras fontes artísticas, e mais recentemente, testes

experimentais de intuições comuns e processos psicológicos subjacentes executados no âmbito da “filosofia experimental”.

Cientistas também têm se interessado pelo estudo dos conceitos populares de emoções, aplicando a eles técnicas experimentais comuns na psicologia dos conceitos. Tais técnicas têm revelado que os conceitos de emoção, como boa parte dos conceitos mais comuns, são prototipicamente organizados (FEHR; RUSSEL, 1984). Há melhores e piores exemplos de emoções comumente entendidas (por exemplo: o medo é um melhor exemplo de emoção do que o assombro) e há casos limitrofes, como o tédio: acerca destes, os usuários da linguagem comum dividem-se quanto a tratarem-se ou não de emoções. Uma variedade de estruturas psicológicas tem sido proposta por teóricos dos conceitos para dar conta da filiação às categorias populares de emoção, incluindo similaridade a protótipos, exemplares, símbolos perceptivos e outros (FEHR; RUSSELL, 1984; WILSON-MENDENHALL *et al.*, 2011).

O que os filósofos e cientistas dos afetos buscam oferecer são definições prescritivas das emoções que sejam maximamente compatíveis com a linguagem comum e com os interesses relativos a objetivos teóricos. Uma razão pela qual os teóricos não estão meramente interessados em esboçar os contornos dos conceitos populares de emoção através de definições descritivas é que eles suspeitam que tais conceitos possam incluir itens muitíssimo diversos a ponto de não serem suscetíveis a quaisquer generalizações teóricas robustas.

À primeira vista, as coisas que mais comumente chamamos de emoções diferem-se uma das outras em vários aspectos. Por exemplo, algumas emoções são ocorrências (como o pânico), e outras são disposições (como a hostilidade); algumas têm vida curta (como a raiva) e outras têm vida longa (como o luto); algumas envolvem processamento cognitivo primitivo (como o medo de um objeto se aproximando repentinamente) e outras envolvem um processamento cognitivo sofisticado (como o medo de perder uma partida de xadrez); algumas são conscientes (como o nojo de um inseto na boca) e outras são inconscientes (como o medo inconsciente de fracassar na vida); algumas têm expressões faciais prototípicas (como a surpresa) e outras não (como o arrependimento). Algumas envolvem fortes motivações para agir (como a ira) e outras não (como a tristeza). Algumas estão presentes em diferentes espécies (como o medo) e outras são exclusivamente humanas (como a *schadenfreude*<sup>36</sup>). E a lista não para por aí.

---

<sup>36</sup> N.T.: Em Alemão: “alegria ao dano”, literalmente. É o regozijo pelo infortúnio alheio. Pode ser íntimo (saboreado pela inveja) ou público (escárnio). É ilustrado na língua portuguesa pela expressão “bem-feito (para ele/a)!”.

Essa heterogeneidade multifacetada levou alguns a concluir que as categorias populares de emoção não designam tipos naturais, seja com respeito à categoria mais genérica de emoção (RORTY, 1980B, 2003; GRIFFITHS, 1997; RUSSELL, 2003; ZACHAR, 206; KAGAN, 2007, 2010), seja com respeito a categorias específicas de emoções, como a raiva, o medo, a felicidade, o nojo e outras (SCARANTINO, 2012; BARRET, 2005, 2017). Não obstante, outros defendem que há, sim, suficiente homogeneidade entre as instâncias de categorias populares de emoção para as qualificarmos como tipos naturais (CHARLAND, 2003; PRINZ, 2004; ZINC, NEWEN, 2008).

Já que o conceito de *tipo natural* é ele próprio controverso e provavelmente mais apto às discussões sobre categorias pelas quais os cientistas dos afetos se interessam, falaremos, alternativamente, de *tipos teóricos*, entendidos como agrupamentos de entidades que participam de um corpo relevante de generalizações filosóficas ou científicas devido a algum conjunto de propriedades que elas tenham em comum.

Que as categorias populares de emoção sejam suficientemente homogêneas a ponto de contarem como tipos teóricos é algo que tem importantes implicações metodológicas. Na medida em que elas sejam homogêneas, as definições prescritivas das emoções que o teórico vier a oferecer podem alcançar tanto a fecundidade teórica quanto a máxima compatibilidade com o uso comum da linguagem (nesse caso, as definições prescritivas também serão descritivamente adequadas). Na medida em que elas não sejam suficientemente homogêneas, as definições prescritivas deverão **explicar** as categorias populares de emoção, transformando-as de modo a aumentar a sua fecundidade teórica ao mesmo tempo que abandonando, em algum grau, a compatibilidade com a linguagem comum (CARNAP, 1950).

A fecundidade teórica, no entanto, é diferentemente concebida pelos filósofos e cientistas dos afetos. Os primeiros costumam ter em vista a compreensão da experiência humana das emoções e por vezes a contribuição com outros projetos filosóficos, tais como explicar as origens da ação racional ou do julgamento moral, lançar luz sobre o que torna a vida digna de ser vivida, ou investigar a natureza do autoconhecimento. Por sua vez, os cientistas dos afetos são mais favoráveis a uma abordagem em terceira pessoa que possa rever muito de nossa autocompreensão em primeira pessoa. E suas definições prescritivas são frequentemente elaboradas para promover a medição e a experimentação com fins de predição e explicação em uma disciplina científica específica.

Neste verbete avaliaremos definições filosóficas e científicas das emoções tanto em termos de sua compatibilidade com a linguagem comum quanto em termos



de sua fecundidade teórica, mas reconhecendo que atualmente faltam, à área, diretrizes que permitam chegarmos ao devido equilíbrio entre esses dois *desiderata*.

## 2. Três tradições no estudo das emoções: emoções como sentimentos, avaliações e motivações

“Emoção” é um termo que veio a ser usado na língua inglesa no século XVII e XVIII como uma tradução do termo francês *émotion*, mas que não designou “uma categoria de estados mentais que pode ser sistematicamente estudada” até a metade do século XIX (DIXON, 2012, p. 338; DIXON, 2003; SOLOMON, 2008). Mas também é o caso que muitas das coisas que nós hoje chamamos de emoções têm sido objeto de análise teórica desde a Grécia Antiga, sob uma variedade de rótulos linguísticos específicos como **paixão**, **sentimento**, **afeto**, **agitação**, **movimento**, **perturbação**, **comoção** ou **apetite**. Daí, tem-se uma longa e complicada história, que, progressivamente, levou ao desenvolvimento de uma variedade de ideias compartilhadas sobre a natureza e as funções das emoções, mas a nenhuma definição consensual do que as emoções são, nem na filosofia, nem na ciência dos afetos.

Uma ideia amplamente compartilhada é a de que as emoções têm componentes, e que tais componentes são conjuntamente instanciados em episódios prototípicos de emoções. Considere um episódio de intenso medo devido à repentina aparição de um urso cinzento em seu caminho durante uma trilha. À primeira vista, podemos distinguir no evento complexo que é o medo um componente **avaliativo** (por exemplo, avaliar o urso como perigoso), um componente **fisiológico** (por exemplo, aumento da frequência cardíaca e da pressão sanguínea), um componente **fenomenológico** (por exemplo, uma sensação desagradável), um componente **expressivo** (por exemplo, sobrancelhas erguidas, mandíbula aberta, lábios esticados horizontalmente), um componente **comportamental** (por exemplo, uma tendência a fugir), e um componente **mental** (por exemplo, focalização da atenção).

Uma questão que tem dividido os teóricos da emoção é: qual subconjunto dos componentes avaliativo, fisiológico, fenomenológico, expressivo, comportamental, e mental é **essencial** à emoção? A resposta a esse “problema das partes” (PRINZ, 2004) tem mudado de tempos em tempos na história do tema, levando a uma vasta coleção de teorias das emoções tanto na filosofia quanto na ciência dos afetos. Ainda que tais teorias diverjam em múltiplos aspectos, elas podem ser categorizadas, com alguma utilidade, em três grandes tradições, que chamaremos de *Tradição*

*sentimentalista, Tradição avaliativa e Tradição motivacional* (SCARANTINO, 2016).

A tradição sentimentalista considera que a maneira pela qual sentimos as emoções é a sua característica mais essencial, e define as emoções como certas experiências conscientes. A tradição avaliativa considera como mais importante o jeito pelo qual as emoções interpretam o mundo, e define as emoções como sendo (ou envolvendo) certas avaliações das respectivas circunstâncias provocantes. A tradição motivacional define as emoções como certos estados motivacionais.

Cada tradição enfrenta a tarefa de articular uma definição prescritiva das emoções que seja teoricamente fecunda e compatível, ao menos em algum grau, com o uso comum da língua. E ainda que haja objetivos teóricos específicos a cada disciplina, também há um núcleo duro de desafios explicativos que tende a ser compartilhado por tais disciplinas:

- *Diferenciação*: Como as emoções diferem-se umas das outras, e das coisas que não são emoções?
- *Motivação*: As emoções motivam comportamentos, e, se sim, como?
- *Intencionalidade*: As emoções são direcionadas a objetos, e, se sim, podem ser adequadas ou inadequadas em relação a seus objetos?
- *Fenomenologia*: As emoções sempre envolvem experiências subjetivas, e, se sim, de que tipo?

Por exemplo, uma explicação adequada da raiva deveria nos dizer como a raiva difere do medo e de estados não-emocionais (diferenciação), se e como a raiva motiva comportamentos agressivos (motivação), se e como a raiva pode ser sobre um dado estado de coisas e se pode ser considerada apropriada a respeito desse estado de coisas (intencionalidade), e se e como a raiva envolve uma experiência subjetiva característica (fenomenologia).

Passaremos agora a tratar de algumas das mais proeminentes teorias em cada tradição, e avaliar como elas se saem nesses quatro desafios teóricos – e outros. Como veremos, cada tradição parece capturar algo de importante sobre como as emoções são, mas nenhuma é imune a contraexemplos e casos problemáticos. Um resultado disso é que a mais recente tendência na teoria das emoções é representada por teorias, tentando combinar ideias de diferentes tradições. Ainda que comecemos nossa investigação com William James e ocasionalmente

mencionemos algumas abordagens anteriores, o nosso principal foco será as teorias desenvolvidas nos últimos 50 anos.

### 3. A tradição sentimentalista inicial: emoções como sentimentos

A mais simples das teorias das emoções, e talvez a que mais represente o senso comum, é a de que as emoções são simplesmente um conjunto de sensações [*feelings*], diferenciados de outras experiências sensoriais (como comer chocolate) ou propriocepções (como sentir dor nas costas) por sua qualidade experienciada. A ideia de que as emoções são um tipo específico de experiências subjetivas tem dominado a teoria das emoções mais ou menos desde a Grécia Antiga até o começo do século XX.

Essa ideia pode ser interpretada de duas formas. Todos os grandes filósofos clássicos – Platão, Aristóteles, Spinoza, Descartes, Hobbes, Hume, Locke – entendiam que as emoções envolviam sensações entendidas como primitivas e sem partes componentes. Uma ideia alternativa foi pela primeira vez introduzida por William James, que defendeu que a psicologia científica deveria parar de tratar os sentimentos [*feelings*]<sup>37</sup> como "entidades psíquicas eternas e sagradas, como as velhas espécies imutáveis na história natural" (JAMES, 1980, p. 449).

A proposta de James, normalmente conhecida como teoria James-Lange por causa de sua notável similaridade com aquela oferecida, na mesma época, por Lange (1885), dizia que as emoções são sentimentos constituídos pelas percepções de mudanças nas condições fisiológicas relativas a funções motoras e autonômicas. Quando percebemos que estamos em perigo, por exemplo, essa percepção provoca diretamente um conjunto de respostas corporais, e nossa consciência de tais respostas é o que constitui o medo. James, assim, defendia que "*a nossa percepção das mudanças [corporais] enquanto elas ocorrem é a emoção*"<sup>38</sup> (JAMES, 1884, p. 189-190).

---

<sup>37</sup> N.T.: O leitor deve notar que traduzimos *feeling* tanto por "sensação" quanto por "sentimento". O assunto é complexo, mas no que tange a essa tradução, ambos os termos se pretendem intercambiáveis, já que são eventos que nos acompanham quando vemos, cheiramos, ouvimos etc.. A diferença, *grosso modo*, é que as sensações são mais facilmente reconhecidas como apenas corporais (sede, cócega, dor, náusea etc.).

<sup>38</sup> N.T.: Ênfase e acréscimo no original.

A teoria James-Lange se sai bem com respeito ao problema da fenomenologia, já que substitui a bruta fenomenologia preferida por explicações anteriores por uma explicação construtivista dos "processos que geram e constroem [...] as experiências conscientes" (MANDLER, 1990, p. 180). Essa abordagem ganhou nova proeminência em tempos mais recentes com a afirmação do movimento *construtivista psicológico* na ciência dos afetos (seção 8.2).

Mas a teoria James-Lange parece ter menos êxito no tocante aos desafios da motivação, diferenciação e intencionalidade. Primeiro, James dizia que o senso comum está errado sobre a direção da causação das emoções e mudanças corporais: uma afirmação mais correta seria a de que

nós sentimos tristeza porque choramos, raiva porque batemos, medo porque trememos, e não choramos, batemos ou trememos porque estamos tristes, raivosos ou com medo. (JAMES, 1884, p. 190)

A implicação contraintuitiva de que as emoções não causam suas manifestações, mas, em vez disso, emergem delas, soa a muitos como problemática porque parece contrariar a ideia de que as emoções são importantes para nós. Como elas poderiam ser tão importantes, perguntaram críticos como Dewey (1894, 1895), se elas não têm papel causal sobre as ações? E por que, poderíamos acrescentar, a ciência não tenta primeiro explicar a causa e a função de tais "estados corporais" originais? E, em especial, **por que** a emoção é sequer provocada, em primeiro lugar? (ARNOLD, 1960).

Além disso, a teoria não explica adequadamente as diferenças entre emoções. Essa objeção foi apresentada, com muita influência, por Walter Cannon (1929). De acordo com uma interpretação comum da teoria James-Lange, o que distingue as emoções entre si é o fato de que cada uma envolve a percepção de um conjunto diferente de mudanças corporais. Cannon respondeu que as reações viscerais características de diferentes emoções, como o medo e a raiva, são indistinguíveis, e que por isso tais reações não seriam aquilo que nos permite apontar as diferenças entre as emoções.

Pesquisas posteriores ainda não concluíram se as emoções têm, de fato, diferentes perfis corporais – a nível neural, expressivo ou autonômico (*vide* CLARK-POLNER *et al.*, 2016; DURAN *et al.*, 2017; KRAGEL; LABAR, 2016; NUMMENMAA; SAARIMÄKI, no prelo; KELTNER *et al.*, 2016). Independentemente de como se

encerre o debate empírico sobre os marcadores corporais, as mudanças corporais ou cerebrais, bem como as sensações que acompanham essas mudanças podem no máximo nos levar até a metade do caminho na direção de uma taxonomia adequada.

Um outro obstáculo de peso frente a teoria James-Lange é que ela não nos dá nenhum *insight* sobre o papel das emoções na nossa vida de agentes racionais pensantes. Mas as emoções podem ser não só explicadas, como também justificadas. Se alguém desperta minha raiva, posso citar o tom depreciativo do meu antagonista; se alguém desperta inveja em mim, posso apontar para a sua transgressão emocional (TAYLOR, 1975). Se as emoções fossem **meras** sensações, como James sugeriu, seria difícil explicar como elas podem ser justificadas por razões, assim como dificilmente nos cabe justificar a experiência sensória de degustar chocolate ou sentir dor nas costas.

#### 4. Emoções e objetos intencionais

Lançar luz sobre o sentido em que as emoções podem ser justificadas exige um breve desvio em direção ao tópico de sua "direcionalidade a objetos", "sobriedade" [*aboutness*] ou "intencionalidade". A primeira distinção que precisamos fazer é a entre objetos **particulares e formais** das emoções. Como Kenny (1963) enfatizou primeiramente, qualquer **X sobre** o qual eu possa ter uma emoção *E* é um objeto particular de *E*, enquanto o objeto formal de *E* é a propriedade que eu implicitamente atribuo a *X* em virtude de ter *E* **a respeito de X**.

Por exemplo, o objeto particular do medo é qualquer coisa que uma pessoa possa temer, enquanto o objeto formal do medo é "aquilo que constitui um perigo", na suposição de que apenas o que é considerado perigoso pode inteligivelmente ser temido. Objetos particulares e formais constituem os dois principais aspectos da intencionalidade emocional: emoções são direcionadas a objetos na medida em que elas tenham objetos particulares, e elas são **adequadas** na medida em que seus objetos particulares instanciem os objetos formais representados pela emoção (*vide* seção 10.1).

A segunda distinção que precisamos fazer é aquela entre dois tipos de objetos particulares das emoções: objetos-alvo e objetos proposicionais (DE SOUSA, 1987). O objeto-alvo de uma emoção é a entidade específica sobre a qual a emoção é. Por exemplo, o amor pode ser sobre Maria, ou sobre Banguécoque, ou sobre Homer Simpson e assim por diante. Todos esses são possíveis alvos do amor, e eles podem ser reais ou imaginários.

Nem toda emoção tem um alvo. Eu posso ficar bravo por minha mulher ter se arrumado de um certo jeito, sem que haja qualquer entidade particular – eu mesmo ou outra pessoa – à qual minha raiva seja direcionada. Inversamente, nem todas as emoções têm um objeto proposicional. Por exemplo, se Maria é o alvo de meu amor, pode não haver proposição alguma, por mais complexa que seja, que capture o que é que eu amo em Maria (KRAUT, 1986; RORTY, 1987 [1988]).

Por fim, também parece haver estados afetivos que são desprovidos de ambos os tipos de objeto particular: eles não são nem direcionados a uma entidade particular, nem são sobre um estado de coisas capturado por uma proposição. Por exemplo, eu posso estar deprimido ou alegre, mas não posso estar deprimido ou alegre a respeito de um alvo ou fato específico qualquer. Esses estados afetivos aparentemente desprovidos de objetos compartilham muitas propriedades com as emoções direcionadas a objetos, especialmente no que diz respeito a seus aspectos fisiológicos e motivacionais, no que podemos, sim, considerá-las como emoções sem objetos.

Por outro lado, alguns têm sugerido que tais estados desprovidos de objetos são melhor encarados como humores [*moods*] (FRIJDA, 1994; STEPHAN, 2017a). Quer pensemos os estados afetivos aparentemente desprovidos de objetos como emoções ou como humores, devemos decidir que tipo de objetos lhes falta. E aqui há duas opções: a primeira é afirmar que alguns estados afetivos não possuem nem objetos particulares nem objetos formais. Se nós pensarmos assim os humores e as emoções desprovidas de objetos, tornar-se-á mais difícil explicar como tais estados afetivos podem ter condições de correção – já que objetos formais são, dentre outras coisas, descrições de como o mundo deve ser para que um estado afetivo seja adequado (TERONI, 2007).

Mas se em vez disso pensarmos tais estados afetivos como tendo objetos formais e condições de correção, então a sua falta de objeto será apenas aparente, porque eles precisariam ter alvos ou objetos proposicionais de algum tipo ao qual implicitamente atribuiríamos a propriedade definida pelo objeto formal. Essa é abordagem sobre humores defendida, dentre outros, por Goldie (2000), que crê que os humores tomam o mundo inteiro como seu objeto, e por Price (2006), que acredita que os humores têm objetos genéricos, mas "ficam atentos" aos particulares.

O que são os objetos formais de emoções específicas? Esse é um assunto controverso, já que a atribuição de objetos formais nos compromete com a tese de que cada emoção, numa base conceitual, atribui uma propriedade específica a seu objeto particular. Isso é frequentemente identificado com um dos vários "temas relacionais centrais" originalmente apresentados por Richard Lazarus (1991, 1991a,

1991b) para explicar que tipos de avaliações causam emoções, uma das principais preocupações da teoria da apreciação na psicologia (seção 6).

Nesse sentido, a raiva representaria desprezos, o medo representaria perigos, a vergonha representaria falhas de viver de acordo com um ego ideal, a tristeza representaria perdas, a felicidade representaria os progressos em direção à realização de um objetivo, o orgulho representaria a melhoria da identidade do ego de alguém (PRINZ, 2004; LAZARUS, 1991b). Uma vez que o objeto formal de uma emoção tenha sido identificado, nós podemos usá-lo para justificar emoções citando suas condições de ocorrência. Por exemplo, se a raiva representa desprezos, então o tom depreciativo de meu antagonista pode ser citado como uma justificativa de minha raiva, porque o tom depreciativo instancia a propriedade, mesma, que a raiva representa.

## 5. A tradição avaliativa inicial na filosofia: emoções como julgamentos

Teorias avaliativas das emoções, também conhecidas como teorias cognitivas das emoções, tornaram-se conhecidas tanto na filosofia como na ciência dos afetos por volta dos anos 1960 e vêm em várias sabores. Uma distinção-chave é a distinção entre teorias avaliativas **constitutivas** e **causais**. Teorias constitutivas afirmam que as emoções são cognições ou avaliações de tipos particulares, enquanto as teorias causais afirmam que as emoções são **causadas por** cognições ou avaliações de tipos particulares. A abordagem constitutiva tende a ser dominante na filosofia, enquanto a abordagem causal goza de significativo apoio na psicologia. Passemos a considerar essas duas correntes do cognitivismo.

A emergência da abordagem constitutiva na filosofia, na metade do século XX, pode ser localizada em um par de artigos por C. D. Broad (1954) e Errol Bedford (1957), e em um livro por Anthony Kenny (1964) (ver também PITCHER, 1965; THALBERG, 1964). Esses autores não foram os primeiros a enfatizar que as emoções são direcionadas a objetos ou dotadas de intencionalidade – Brentano (1874 [1995] já havia feito isso com a inspiração de vários autores medievais (KING, 2010). Mas esses filósofos da metade do século XX foram os primeiros a articular uma influente teoria de que, para podermos explicar sua intencionalidade, as emoções deveriam ser avaliações cognitivas de um certo tipo, em vez de sensações [*feelings*] (*vide* MEINONG, 1894).

O argumento é em resumo o seguinte: se emoções têm intencionalidade, segue-se que há padrões internos de adequação segundo os quais uma emoção é adequada apenas no caso de seu objeto formal ser instanciado (KENNY, 1963). Mas sensações [*feelings*] não são os tipos de coisas que podem entrar em relações conceituais com objetos formais. Logo, para poderem estar corretamente ancoradas em relações conceituais desse tipo, as emoções devem ser ou envolver "avaliações cognitivas" de algum tipo.

Que tipos de avaliações cognitivas? A teoria cognitivista mais parcimoniosa segue os estoícos em caracterizar as emoções como julgamentos. Robert Solomon (1980), Jerome Neu (2000) e Martha Nussbaum (2001) adotam essa abordagem. Em uma interpretação comum dessas teses, minha raiva de alguém é o julgamento de que eu fui prejudicado por essa pessoa. Generalizando-a, a ideia é que uma emoção *E* é um julgamento de que o objeto formal de *E* foi instanciado (por um objeto particular *X*).

A teoria é normalmente referida como *judgamentalismo* [*judgementalism*], mas o rótulo é potencialmente enganador, pois sugere que para os proponentes da teoria uma emoção nada mais é do que um julgamento, entendido como um assentimento a uma proposição. Essa interpretação é de fato pressuposta por alguns de seus críticos corriqueiros.

Em primeiro lugar, é dito que o judgamentalismo não explica como as emoções podem motivar, pois podemos fazer um juízo – digamos, o juízo de que fui prejudicado – sem que eu seja motivado a agir a respeito disso. Em segundo lugar, ela não explica a fenomenologia das emoções, já que a um julgamento faltam as dimensões corpórea, de valência e suscitação que tipicamente caracterizam a experiência da emoção. Em terceiro lugar, ela falha em explicar as emoções de animais e bebês, que supostamente não possuem a capacidade de assentir a proposições (DEIGH, 1994). Em quarto lugar, ela não explica a "resistência à razão" que algumas emoções possuem quando não são extinguidas por julgamentos que as contradizem, como quando alguém julga que voar de avião não é perigoso, mas continua a ter medo disso (D'ARMS; JACOBS, 2003).

Os judgamentalistas têm buscado responder a essas críticas esclarecendo que tipo de julgamentos as emoções são (e alguns, como Nussbaum e Neu, têm explicitamente rejeitado o rótulo de judgamentalismo). Tem sido dito, por exemplo, que nós deveríamos pensar os julgamentos como "envolvendo um desejo central" (SOLOMON, 2003, p. 105-6), o que os tornaria motivacionais (por exemplo: o medo envolve o desejo central de fugir). Tais julgamentos também seriam



"dinâmicos" e capazes de "abrigar [...] os movimentos de desordem da emoção" (NUSSBAUM, 2001, p. 45), sendo assim fenomenologicamente salientes. Já que envolvem a aceitação pré-linguística e não-linguística de como o mundo se parece, eles estariam disponíveis a bebês e animais. E, por fim, sendo capazes de estar par a par com julgamentos contraditórios, eles poderiam explicar a resistência à razão (NUSSBAUM, 2001).

Várias objeções têm sido lançadas contra essa estratégia. Tomando apenas um exemplo proeminente dentre muitos, tem sido dito que explicar a resistência à razão em termos de julgamentos contraditórios – julgar que  $p$  e que não- $p$  – atribui a agentes o tipo errado de racionalidade (HELM, 2001, 2015; DÖRING, 2008; BENBAJI, 2012; BRADY, 2009; TAPPOLET, 2000; FAUCHER; TAPPOLET, 2008a,b). Provavelmente também há um problema maior aqui: o de que os julgamentalistas ampliam demais o conceito de "julgamento", sem princípio algum, de modo a dar conta de todos os contraexemplos, em vez de distinguir os tipos significativamente diferentes de estados cognitivos, todos agrupados pelo mesmo nome.

O problema dessa **estratégia elástica** não é apenas o de que ela é *ad hoc*, mas também o de que ela nos mete em uma conversa confusa, repleta de mal-entendidos e em última instância equivale a uma "vitória de Pirro" da teoria avaliativa, já que, em uma acepção suficientemente expandida da noção de julgamento, a identificação das emoções com julgamentos torna-se, na melhor das hipóteses, trivialmente verdadeira e fracassa em lançar luz sobre o que as emoções de fato são (SCARANTINO, 2010).

Dois estratégias mais promissoras têm sido lançadas para defender o cognitivismo de contraexemplos. A primeira, que podemos chamar de *estratégia julgamentalista de adição* (GOLDIE, 2000), consiste em explicitamente acrescentar aos julgamentos outros componentes das emoções, em vez de incorporá-los em julgamentos através da estratégia elástica. Por exemplo: a dimensão motivacional das emoções tem sido explicada pela sugestão de que as emoções não são julgamentos, mas, na verdade, combinações de julgamentos (ou crenças) com desejos (MARKS, 1982; GREEN, 1992; GORDON, 1987). Outros autores têm incorporado mais elementos, propondo que as emoções são um misto de julgamentos, desejos e sensações, em um movimento destinado a explicar tanto a dimensão motivacional quanto a fenomenológica das emoções (LYONS, 1980).

Outra estratégia, que podemos chamar de *estratégia das cognições alternativas*, consiste em substituir a noção ampla de julgamento por uma variedade de outros tipos de avaliações cognitivas que poderiam explicar a intencionalidade

das emoções, ao passo que evitariam algumas das críticas que têm sido levantadas contra o julgamentalismo. Uma vez que muito do que a filosofia das emoções contemporânea tem feito é focar em quais das cognições alternativas devem ser preferidas, dedicaremos uma seção inteira a esse assunto. Primeiramente, discutimos como a Tradição Avaliativa foi desenvolvida na ciência dos afetos.

## 6. A tradição avaliativa nas ciências afetivas: teorias da apreciação

Por volta da mesma época em que a Tradição Avaliativa se tornou popular na Filosofia, uma tradição paralela emergiu na ciência dos afetos através da obra pioneira de Magda Arnold e Richard Lazarus. O que impulsionou esse desenvolvimento foi em parte a revolução cognitivista, o movimento intelectual que substituiu o comportamentalismo nos anos 1960 e colocou o processamento cognitivo de representações mentais no centro da ciência da psicologia.

Arnold defendeu que a pesquisa sobre as emoções havia recusado explicar como as emoções são provocadas. Para lançar luz sobre esse ponto, ela introduziu a noção de *apreciação* [*appraisal*], o processo que determina o significado de uma situação para um indivíduo. A apreciação dá origem à atração ou à aversão, e para Arnold a emoção equivale a essa

inclinação sentida para perto de qualquer coisa intuitivamente apreciada como boa (benéfica), ou para longe de qualquer coisa intuitivamente apreciada como ruim (danosa). (ARNOLD, 1960, p. 171)

Diversos autores anteriores a Arnold haviam reconhecido que as emoções devem ser produzidas por um tipo de avaliação cognitiva das circunstâncias eliciadoras, seja na forma de um julgamento, pensamento, percepção ou ato de imaginação. Afinal de contas, é bastante claro que o mesmo estímulo pode gerar diferentes emoções em diferentes pessoas, ou na mesma pessoa em diferentes momentos, o que sugere que não são os estímulos enquanto tais que provocam as emoções, mas os estímulos enquanto apreciados.

Arnold (1960) foi a primeira a colocar a estrutura interna do processo de apreciação sob escrutínio científico. Apreciações, ela sugeriu, são feitas em três dimensões principais: circunstâncias provocadoras podem ser avaliadas como boas

ou ruins, presentes ou ausentes, e fáceis de se obter ou evitar. Por exemplo: a avaliação cognitiva que causa o medo pode ser descrita como uma apreciação de um evento como ruim, ausente, mas possível no futuro, e difícil de se evitar; enquanto a causa da alegria poderia ser descrita como a apreciação de um evento como bom, presente e fácil de se manter.

Em geral, as teorias da emoção como apreciação são explicações da estrutura dos processos que extraem significado dos estímulos e diferenciam as emoções umas das outras. Também se considera, com frequência, que a apreciação é um processo dinâmico: apreciações seriam seguidas de reapreciações, seguindo mudanças no ambiente e em variáveis internas, moldando as emoções incrementalmente ao longo do tempo.

Note-se que as teorias da apreciação não são exatamente teorias sobre o que as emoções são, ainda que teóricos da apreciação tenham, no mais das vezes, buscado articular tais teorias como um complemento às suas teorias da estrutura da apreciação. Mais especificamente, teorias da apreciação são em princípio compatíveis com as teorias das emoções que as identificam como avaliações, sensações ou motivações, pelo menos na medida em que essas teorias reconheçam que as apreciações desempenham um papel essencial na diferenciação das emoções umas das outras. Dito isso, muitos teóricos da apreciação influentes – incluindo Arnold, Lazarus e Scherer – apresentam teorias das emoções que melhor se encaixariam na Tradição Motivacional.

Teorias científicas têm aumentado significativamente nosso entendimento da natureza da apreciação, dotando-a de ainda mais estruturas do que Arnold originalmente previra (C. SMITH; ELLSWORTH, 1975; FRIJDA, 1986; LAZARUS, 1991, A,B; ROSEMAN, 1996; SCHERER, 2001; ELLSWORTH; SCHERER, 2003; ROSEMAN; SMITH, 2001; OATLEY; JOHNSON-LAIRD, 1987).

Lazarus (1991, 1991b) por exemplo, introduziu seis dimensões estruturais de apreciação, incluindo (1) a relevância do objetivo, (2) a congruência ou incongruência do objetivo, (3) o tipo de ego-envolvimento, (4) a culpa ou mérito, (5) o potencial de enfrentamento, e (6) a expectativa de futuro. Por exemplo: considera-se a culpa como causada pela apreciação de um evento como um objetivo relevante, incongruente, envolvendo uma transgressão moral, e pelo qual o *self* deve ser culpado (as apreciações do potencial de enfrentamento e de expectativa de futuro ficam em aberto). A teoria das emoções do próprio Lazarus é rotulada como cognitivo-motivacional-relacional porque defende que

uma emoção é um estado complexo, um AB, com [a apreciação] A como causa de B como uma combinação de uma tendência para a ação, mudança fisiológica e afeto subjetivo, (LAZARUS, 1991a, p. 819)

através da qual a apreciação não é apenas a causa da emoção, mas também é parte dela. Para uma crítica dessa suposição, confira Moors (2013).

Scherer *et al.* (2001) distinguiu dezesseis dimensões da apreciação, chamadas de Critérios de Avaliação de Estímulo [*Stimulus Evaluation Checks*, SECs], que podem ser agrupados em quatro classes: apreciações de relevância, apreciações de consequências, apreciações de potencial de enfrentamento, e apreciações de importância normativa. No modelo de processo componente de Scherer,

a emoção é definida como um episódio de mudanças inter-relacionadas e sincronizadas nos estados de todos ou quase todos os cinco subsistemas orgânicos, em resposta à avaliação de um evento-estímulo externo ou interno como relevante para as preocupações maiores do organismo. (SCHERER, 2005, p. 697)

Os cinco subsistemas orgânicos subjazem a cinco componentes das emoções que, quando envolvidos em mudanças coordenadas, instanciam as emoções: uma apreciação, mudanças fisiológicas autonômicas, uma tendência à ação, uma expressão motora e uma sensação subjetiva. Em sua mais recente colaboração com Moors (MOORS; SCHERER, 2013), Scherer sugeriu que o propósito de cada critério de avaliação de estímulo é determinar tendências à ação que dariam um lugar de destaque ao componente motivacional, tornando a sua teoria um híbrido avaliativo-motivacional (*vide* SCHERER; MOORS, 2019).

Uma variante das teorias da apreciação tem recentemente atraído algum interesse na computação afetiva, uma abordagem interdisciplinar que combina ideias da ciência dos afetos e da ciência da computação (*vide* PICARD, 1997). Falemos da Teoria das Emoções da Crença e do Desejo (TECD) desenvolvida por Reisenzein (2009a, 2009b;). Confira, também, Miceli e Castelfranchi (2015). A TECD afirma que as emoções são causadas por uma combinação de avaliações cognitivas (crenças) e motivações conativas (desejos), enquanto as teorias da apreciação comuns postulam avaliações cognitivas de congruência-motivo, o que de fato avalia

o grau em que o estímulo é congruente com os objetivos/desejos do agente.

A tese central da TECD é a de que as emoções são provocadas por mecanismos inatos, cuja função evolucionária é a de comparar crenças recentemente adquiridas com desejos e crenças existentes, de modo a monitorar e atualizar o sistema central representacional humano (o sistema de crenças e desejos). Por exemplo: suponha que você tenha a crença de que seu candidato favorito perderá a eleição, mas deseja que ele ganhe a eleição (REISENZEIN, 2009a). Assim que a nova informação de que ele de fato ganhou a eleição é obtida, um sistema comparador de crença-com crença produz um sinal de infirmação de crença, subjetivamente experimentado como surpresa, e um sistema comparador de crença com desejo produz um sinal de satisfação de desejo, subjetivamente experimentado como prazer. Isso, por sua vez, leva a respostas adaptativas que incluem uma reorientação da atenção ao novo conteúdo de que seu candidato favorito venceu, a supressão da crença de que seu candidato favorito iria perder, e a experiência subjetiva de felicidade.

Generalizando, para a TECD as emoções são combinações de sinais de infirmação/confirmação de crenças, e sinais de frustração/satisfação de desejos experimentados, respectivamente, como confirmações de surpresa/antecipação e prazer/desprazer, que se misturam na experiência emergente de uma emoção específica (por exemplo: felicidade, medo, esperança etc.). Tais sinais são não-conceituais no sentido de que não pressupõem o uso de conceitos e ocasionam o redirecionamento da atenção, atualizações do repertório de crenças e desejos, e, quando acima de um limiar, experiências subjetivas distintivas.

Um desafio encarado pelas teorias da apreciação diz respeito a se as apreciações são as causas das emoções, consequências das emoções, partes das emoções, ou uma combinação destas. Essas perguntas levantam questões conceituais complexas que não podemos abordar aqui (*vide* MOORS, 2013), mas são essenciais para avaliarmos o grau de apoio que as evidências dão à teoria da apreciação. Uma crítica de longa data a esse programa de pesquisa (OATLEY, 1992; PARKINSON; MANSTEAD, 1992; PARKINSON, 1997; RUSSELL, 1987; FRIJDA, 1993) é a de que as evidências dadas por relatos em primeira pessoa, comumente usadas para apoiar uma interpretação causal da relação entre apreciações e emoções, apenas apoia uma relação de implicação conceitual entre elas, uma vez que revela nossas crenças sobre o que torna as emoções **adequadas** em vez de o que **causa** as emoções (*vide* Scherer, 2009).

## 7. A tradição híbrida sentimentalista-avaliativa na filosofia recente

Mencionamos anteriormente que uma conhecida resposta às críticas recebidas pelo julgamentalismo filosófico tem sido a *estratégia de cognições alternativas*, que pretende explicar melhor a intencionalidade, a diferenciação, o poder motivacional e a fenomenologia, bem como o potencial de resistência das emoções a razões.

Isso levou a uma gradual convergência das tradições sentimentalista e avaliativa, com a primeira passando a identificar as emoções como percepções avaliativas dotadas de uma certa fenomenologia, e a última passando a identificar as emoções como sentimentos avaliativos com uma certa intencionalidade. Como resultado disso, a distinção entre teorias avaliativas (ou cognitivistas) e teorias sentimentalistas borrou-se significativamente, com a maioria das explicações filosóficas das emoções passando a serem híbridas.

### 7.1 Emoções como percepções avaliativas

Teorias perceptuais vem nas variedades literal/forte e não-literal/fraca (BRADY, 2013; SALMELA, 2011). Versões fortes normalmente assumem que as emoções são formas genuínas de percepção, na linha da percepção sensorial. Versões fracas sublinham propriedades-chave que as emoções compartilham com a percepção sensorial, ao mesmo tempo que reconhecem diferenças importantes.

A teoria neojamesiana de Prinz é um bom exemplo de uma teoria perceptiva forte. Para Prinz (2004), podemos falar *bona fide* de um sistema perceptivo quando estamos diante de um sistema de **entradas** dedicado com transdutores especializados e representações mentais. A percepção sensorial claramente tem um sistema de **entradas** dedicado à visão, olfação, audição, tato e paladar. Seguindo os passos da obra neurocientífica de Damasio (1994, 2003), Prinz sugere que as emoções também podem depender de um sistema dedicado dentro do sistema somatossensorial. Assim, as emoções literalmente seriam percepções de mudanças corporais, quer num nível visceral, hormonal ou musculoesquelético, quer na forma de mudanças nas áreas somatossensoriais do cérebro.

Prinz acrescenta que as emoções não são *apenas* percepções de mudanças corporais, do que se segue que duas emoções podem diferenciar-se uma da outra apesar de envolverem percepções de mudanças corporais indistinguíveis. Por

exemplo: o medo não é apenas a percepção de “uma aceleração cardíaca e [...] outras mudanças fisiológicas” (PRINZ, 2004, p. 69): ele também tem a função distintiva – ser provocado pelo perigo – e uma marcador específico de valência negativa que motiva a ação repelente. Já que numa teoria teleossemântica da representação os estados mentais representam o que eles têm a função de indicar (PRINZ, 2004; DRETSKE, 1988), Prinz conclui que perceber o coração acelerar também pode representar perigo desde que ele tenha a função de o indicar (*vide* SHARGEL; PRINZ, 2018; ROBINSON, 2005). Em resumo, sujeitos literalmente percebem mudanças corporais (o conteúdo nominal) e percebem indiretamente o objeto formal (conteúdo real) em virtude do fato de que as mudanças corporais representam objetos formais.

Teorias perceptivas fracas consideram as emoções como significativamente análogas à percepção sensorial ou propriocepção. Além disso, a maioria considera as emoções como sendo percepções diretas de objetos formais, em vez de percepções de mudanças corporais com a função de rastrear objetos formais.

Uma influente proposta nessa linha é apresentada por Roberts (2003), que afirma que “as emoções são um tipo de percepção” (2003, p. 87) na forma de interpretações baseados em interesses. Roberts deixa claro que interpretações são “impressões, jeitos pelos quais as coisas aparecem ao sujeito” (2003, p. 75) e que são baseadas em interesses em virtude de serem baseadas nos desejos e aversões do sujeito. Por exemplo: o medo de um pai que o fogo machuque sua filha é uma interpretação do fogo como perigoso, baseado no desejo do pai de que nada de ruim aconteça à sua filha.

De modo similar, Tappolet (2016) sugere que as emoções são experiências perceptivas de propriedades valorativas (ou seja: valores) como periculosidade (medo) ou desprezo (raiva). Alguns autores acrescentam que tais propriedades valorativas não ficam disponíveis através de quaisquer outros meios, assim como as propriedades cromáticas não nos são disponíveis exceto pelo canal da percepção visual (*vide* JOHNSTON, 2001).

Tappolet enfatiza que a percepção avaliativa, exatamente como a percepção sensorial, é não-conceitual por natureza e é cognitivamente impenetrável (ver também DÖRING, 2007; DÖRING; LUTZ, 2015; GOLDIE, 2000; TAPPOLET, 2000; GOLDIE, 2002; WOLLHEIM, 1999; CHARLAND, 1995; ZAJONC, 2000). Isso explicaria por que criaturas que não possuem conceitos, como os animais e as crianças pré-linguísticas, podem ter emoções, e explicaria a resistência emocional a razões, o que pode ser entendido nos mesmos termos de uma ilusão visual. Assim

como percebemos visualmente um lápis como entortado, ao mesmo tempo que o julgamos como reto, também percebemos emocionalmente uma plataforma transparente sobre o Grand Canyon como perigosa, ao mesmo tempo que a julgamos como não-perigosa.

Tappolet (2016) lista características adicionais que ajudariam a explicar por que tantos autores chegaram a pensar que as emoções são percepções: (a) tanto as emoções quanto as percepções tem propriedades fenomenais salientes, (b) ambas são automaticamente provocadas por objetos reais ou imaginados, (c) ambas têm condições de correção porque representam o mundo como sendo de um certo jeito, e (d) ambas desempenham o papel epistêmico de prover razões revogáveis<sup>39</sup> para uma crença (por exemplo: a percepção visual para a crença de que algo é azul, e o medo para a crença de que algo é perigoso).

Apesar dessas analogias, muitos críticos rejeitaram a teoria perceptiva das emoções (SALMELA, 2011; DOKIC; LEMAIRE, 2013). Uma proeminente crítica refere-se à incapacidade da teoria de explicar a resistência emocional a razões. Por exemplo: Helm (2001) defende que os perceptualistas acabaram removendo a irracionalidade que é distintiva da resistência. Se perceber uma plataforma transparente sobre o Grand Canyon como perigosa ao mesmo tempo que se a julga como não-perigosa fosse exatamente como uma ilusão visual, então não haveria nada de irracional sobre isso, como se não houvesse nada de irracional em ver um lápis como entortado ao mesmo tempo que o julgamos como reto. Mas claramente há uma medida de irracionalidade envolvida em emoções resistentes a razões: ao contrário das ilusões perceptivas, elas nos motivam a agir. Em outras palavras: elas envolvem um assentimento passivo que contradiz o assentimento ativo capturado pelo juízo contraditório.

## 7.2 Emoções como sentimentos avaliativos

Diversos autores têm proposto teorias que dotam os sentimentos de intencionalidade. Um exemplo notável é Goldie (2000), que identifica a intencionalidade das emoções com aquela dos **sentimentos direcionados a**, que não são apenas

---

<sup>39</sup>N.T.: Raciocínio revogável, ou anulável (*defeasible reasoning*) é, segundo esta mesma enciclopédia, o tipo de argumentação racionalmente persuasiva, mas dedutivamente inválida.



sensações corporais que tomam sua intencionalidade de algum outro lugar, como na abordagem de Prinz, mas que em vez disso teriam sua própria intencionalidade (ver também DÖRING, 2007; PUGMIRE, 1998). Por exemplo: quando eu sinto medo de escorregar no gelo, meu sentimento é direcionado ao gelo como sendo perigoso. Esse tipo de sentimento é uma questão de pensar o gelo com sentimento, e não pode ser reduzida à combinação de uma sensação corporal não-intencional e um pensamento avaliativo não-emocional. Como Goldie coloca,

sentimentos emocionais são intrinsecamente ligados ao caráter direcionado ao mundo das emoções, de modo que uma explicação adequada da intencionalidade das emoções (...) ao mesmo tempo irá capturar um importante aspecto de sua fenomenologia. (GOLDIE, 2002, p. 242; *vide* RATCLIFFE, 2005, 2017)

De modo similar, Helm (2009, p. 8) propõe que "as emoções são sentimentos intencionais de importância", prazerosos ou desprazerosos. Segundo Helm,

que uma coisa importe a você – para que você dê *bola* para ela – significa (grosso modo) que ela seja digna de atenção e ação. (HELM, 2009, p. 252, ênfase do original)

Isso explica o porquê que as emoções motivam a ação: o sentir que alguma coisa é digna de atenção e ação é estar motivado. Isso também explica o que torna a resistência emocional a razões irracionais. Ao julgar uma plataforma transparente sobre o Grand Canyon como não-perigosa, e mesmo assim temê-la, o sujeito julga que aquilo que ele ou ela sente como digno de atenção ou ação na verdade não é, assumindo assim compromissos avaliativos em desacordo uns com os outros (HELM, 2015, p. 430-431).

Algumas das abordagens que atribuem intencionalidade aos sentimentos são inspiradas pelo programa de pesquisa mais amplo do *representacionalismo* na filosofia da mente, que é a tese de que as propriedades fenomenais são identificáveis com, ou ao menos redutíveis a, propriedades intencionais (DRETSKE, 1995; HORGAN; TIENSON, 2002). Em algumas variantes do representacionalismo, a fenomenologia emocional que é reduzida é a meramente somática, no sentido de que a sensação é direcionada a eventos corporais (TYE, 1995). Em outras variantes,

a fenomenologia é bastante mais rica, incluindo componentes somáticos, cognitivos, conativos e irredutivelmente afetivos, direcionados a objetos particulares e formais no mundo (KRIEGEL, 2012).

Uma alternativa adotada por alguns teóricos contemporâneos é afirmar que as emoções são sentimentos desprovidos de quaisquer objetos intencionais. A impressão contrária seria uma ilusão originada do fato de que os sentimentos que chamamos de emoções são tipicamente causados por pensamentos com uma estrutura intencional com a qual eles são associados em uma experiência composta (WHITING, 2011; GOLDSTEIN, 2002). Nessa abordagem, o medo do gelo é um estado mental composto consistindo em uma emoção – o sentimento de medo, desprovido de objeto – mais um pensamento que tem o gelo como objeto intencional. Em si mesmas, as emoções são sensações meramente hedônicas sem intencionalidade alguma. No entanto, as razões apresentadas para essa tese são explicitamente fenomenológicas, e uma vez que a introspecção da maioria dos pesquisadores parece dar o veredicto contrário – o de que as emoções são em si mesmas direcionadas a objetos –, a tese da composição fracassa em convencer. Para um argumento diferente visando defender a ausência de objeto, baseado em dados da psicologia social, consulte Shargel (2015).

### 7.3 Emoções como padrões de saliência

Outra influente abordagem na recente filosofia das emoções considera que essas sejam:

mecanismos que controlam os fatores cruciais da saliência dentre aquilo que caso contrário seria uma miríade intratável de objetos de atenção, interpretações, e estratégias de inferência e conduta. (DE SOUSA, 1987, p. XV; *vide* ELGIN, 2006; EVANS, 2001; BEN-ZE'EV, 2000).

Por exemplo, há um sem-número de coisas que eu em princípio poderia focar quando estivesse diante de um urso cinzento durante uma trilha, mas meu medo foca minha atenção exatamente no urso, em como interpretar seus movimentos, e em como inferir e executar uma estratégia de fuga. Essa abordagem pode ser

considerada como tendo um “caráter perceptivo” (PRINZ, 2004, p. 223) já que descreve as emoções como mecanismos de mudança da saliência, e percepções certamente podem afetar a saliência. Mas de Sousa busca chamar a atenção para o papel mais amplo que as emoções desempenham em fornecer uma estrutura para as cognições de ambas as variedades perceptivas (por exemplo, o que vemos e o que ouvimos) e não-perceptivas (como o que acreditamos e o que lembramos).

Alguns filósofos sugerem que o poder diretivo que as emoções exercem sobre as cognições é parcialmente uma função de sua estrutura essencialmente dramática ou narrativa (RORTY, 1986 [1988]). Goldie (2012) oferece uma investigação particularmente sutil do papel da narrativa em constituir nossas emoções ao longo do tempo. Parece conceitualmente incoerente supor que possamos ter uma emoção – digamos, um intenso ciúme ou uma raiva dominante – por apenas um átimo de segundo (WOLLHEIM, 1999). Uma explicação dessa característica das emoções é a de que há uma história a desenrolar-se durante cada episódio emocional, e histórias ocorrem ao longo de períodos de tempo. Curiosamente, Goldie afirma que a estrutura narrativa das emoções é a mesma independentemente de as emoções serem experimentadas em relação a objetos reais ou ficcionais, o que explicaria por que podemos responder a personagens ficcionais com respostas emocionais completas, ainda que mudas do ponto de vista motivacional. Para um levantamento das outras soluções ao chamado *paradoxo da ficção*, consulte Cova e Friend, no prelo.

De Sousa (1987) sugeriu que as histórias características de diferentes emoções são aprendidas por associação com certos “cenários paradigmáticos”. Cenários paradigmáticos envolvem dois aspectos: primeiro, um tipo de emoção que providencia os objetos característicos de um tipo específico de emoção (em que os objetos podem ser particulares e formais), e segundo, um conjunto de respostas características ou “normais” à situação, em que a normalidade é determinada por um misto complexo e controverso de fatores biológicos e culturais. Esses cenários são tirados primeiramente de nossa vida cotidiana quando pequenos, e mais tarde reforçados pelas histórias, arte e cultura a que somos expostos. Ainda mais tardiamente, eles são complementados e refinados pela literatura e outras formas de arte capazes de expandir os limites da imaginação de alguém sobre os modos de se viver (DE SOUSA, 1990; FAUCHER; TAPPOLET, 2008b).

Uma vez que nosso repertório emocional fica estabelecido, passamos a interpretar novas situações através das lentes de diferentes cenários paradigmáticos. Quando um cenário específico se insinua como uma interpretação, ele combina ou recombina nossas disposições inferenciais, cognitivas e perceptivas. Quando um

cenário paradigmático é suscitado pela situação nova, a emoção resultante pode ou não ser adequada à situação que a dispara. Assim, o medo de palhaços da infância pode ser reavaliado e superado na vida adulta como um resultado de uma avaliação mais realista. Ao menos nesse sentido, as emoções podem ser avaliadas racionalmente (*vide* seção 10 para uma maior discussão).

## 8. A tradição motivacional nas ciências afetivas e seus oponentes

A terceira tradição no estudo das emoções as caracteriza como sendo essencialmente tipos de estados motivacionais – entendendo esses, em resumo, como uma causa interna de comportamentos dirigidos à realização de um objetivo. Membros dessa tradição de pesquisa acreditam que o problema central da teoria das emoções precisa ser o de chegar a uma explicação para como as emoções e as ações se relacionam, já que é em última instância o que fazemos quando nos emocionamos que gera relevância pessoal e consequências sociais.

A teoria motivacional foi antecipada por muitos teóricos das emoções da Grécia Antiga e do período da Idade Média que enfatizavam a relação constitutiva entre as emoções e os impulsos comportamentais (KING, 2010), mas teve em Dewey (1894, 1895) seu primeiro precursor moderno. Dewey estava insatisfeito com a inversão do senso comum proposta pela ideia Jamesiana de que as emoções são sensações que emergem em resposta a propriocepções. Se nós realmente ficamos bravos **porque** batemos e mordemos, Dewey objetou, então a raiva não pode causar a batida e a mordida, e isso privaria a raiva, assim como outras emoções, de sua importância explicativa.

A principal sugestão de Dewey era a de que há uma diferença entre a sensação da raiva e a raiva em si: uma emoção “em sua inteireza” é “um modo de comportamento que é intencional [*purposive*]” e “que também se reflete em sensações” (DEWEY, 1895, p. 15). Quando dizemos que alguém está com raiva, Dewey conclui, “nós não apenas, ou nem sobretudo, queremos dizer que [tal pessoa] ‘sente’ uma certa coisa a ocupar sua consciência”. Na verdade, “queremos dizer que ele [...] assumiu uma prontidão para agir de certas maneiras” (DEWEY, 1895, p. 16-17). A ideia de que as próprias emoções sejam ou mecanismos que mudam nossa prontidão para agir, ou os próprios estados de prontidão para a ação foi desde então desenvolvida de várias formas nas ciências afetivas e na filosofia das emoções. Começemos por uma variante evolucionária influente da Tradição Motivacional nas ciências afetivas.

## 8.1 Teoria básica da emoção: emoções como programas de afeto evoluídos

A teoria básica da emoção (TBE) emergiu nos anos 1970 da pioneira obra de Silvan Tomkins, cuja ideia principal era o de que “o sistema motivacional primário é o sistema afetivo” (TOMKINS, 2008, p. 4) Tomkins propôs que há nove afetos básicos ou inatos controlados por programas herdados: interesse, gozo, surpresa, medo, raiva, aflição, vergonha, satisfação e nojo. Seu poder motivacional vem de os sentirmos como prazerosos ou não prazerosos, emergindo da percepção de mudanças faciais que providenciam “informação [feedback] motivacional” (TOMKINS, 2008, p. 623).

A teoria básica dos afetos de Tomkins foi seguida por dois desenvolvimentos relacionados. O primeiro foi o nascimento da teoria básica moderna da emoção, com sua atenção total à universalidade das expressões faciais, presente sobretudo na obra de Paul Ekman (EKMAN *et al.*, 1972; EKMAN, 1980, 1999A, 2003; EKMAN; FRIESEN, 1969) e Carrol Izard (1971, 1977, 1992, 2007). O segundo foi a emergência da abordagem da psicologia evolutiva às emoções, entendidas como soluções a problemas evolutivos recorrentes, e de contribuições proeminentes por Plutchik (1980), e Tooby e Cosmides (2008). Para os primeiros exemplos de teorias evolucionárias das emoções, confira, também, Shand (1920) e McDougall (1908 [2001]).

Pelos anos 1990 as duas abordagens começaram a se juntar, ainda que os psicólogos evolutivos, mais do que os teóricos básicos da emoção, tendam a dizer que uma dada emoção resolve um problema evolutivo apenas com base em argumentos de plausibilidade. Segundo Ekman (1999a, p. 46) “as emoções evoluíram por seu valor adaptativo em lidar com as tarefas fundamentais da vida”, como

lutar, apaixonar-se, fugir de predador, enfrentar a infidelidade sexual, experimentar uma perda de status causada pelo fracasso, responder à morte de um membro da família. (TOOBY; COSMIDES, 2007, p. 117; *vide* KELTNER; HAIDT, 2001).

Elas têm valor adaptativo porque elas rapidamente mobilizam e coordenam os recursos necessários para termos sucesso em lidar com tais tarefas, e porque elas comunicam informações socialmente relevantes via expressões corporais. Assim que um programa básico de emoção é ativado,

uma série de mudanças (com ou sem a consciência imediata) ocorre em instantes de segundos: nos sinais emocionais da face e voz; nas ações pré-determinadas; nas ações aprendidas; no sistema nervoso autônomo que regula nossa corpo; nos padrões regulatórios que continuamente modificam nosso comportamento; no resgate de expectativas e memórias relevantes, e em como interpretamos o que está acontecendo dentro de nós e no mundo. (EKMAN; CORDARO, 2011, p. 366)

Disso se segue a hipótese empírica central da TBE tradicional: deve haver assinaturas corporais para cada emoção básica, consistindo em mudanças altamente correlatas e específicas a cada uma das emoções, no nível das expressões faciais, das mudanças autônomas e das ações pré-determinadas e adquiridas. Mais especificamente, Ekman definiu as emoções básicas em termos de (a) sinais universais distintivos, (b) fisiologia distintiva, (c) pensamentos, memórias e imagens distintivos, (d) experiências subjetivas distintivas, (e) aparência de desenvolvimento previsível, (f) presença homóloga em outros primatas, (g) apreciações automáticas afinadas a universais distintivos em eventos antecedentes, (h) início rápido, com curta duração e ocorrência espontânea (EKMAN, 1999a, p. 5). Alguns teóricos da emoção básica também sugerem que as emoções básicas são associadas com circuitos neurais inatos distintivos (IZARD, 2011; LEVENSON, 2011; PANKSEPP, 1998, 2000).

Munido dessa definição, Ekman passa a afirmar que temos evidência empírica para seis programas afetivos básicos (felicidade, tristeza, raiva, medo, nojo e surpresa), para então expandir a lista de modo a incluir estados cujo estatuto básico seria provavelmente demonstrado no futuro, como a diversão, o desprezo, o acanhamento, a animação, a culpa, o orgulho pelo sucesso, o alívio, a satisfação, o encantamento, o êxtase, o prazer sensorial e a vergonha (EKMAN; CORDERO, 2011). Filósofos de linha cientificista costumam restringir suas discussões sobre as emoções a programas afetivos básicos, já que afirmam que esses são os únicos tipos naturais até agora descobertos no campo dos afetos (GRIFFITHS, 1997; DELANCEY, 2002).

A principal fonte de evidências para os programas afetivos básicos vem de estudos transculturais sobre as expressões faciais que utilizam uma técnica de

reconhecimento descrita pela primeira vez por Darwin (1872). Ela consiste em mostrar imagens de expressões emocionais e perguntar aos observadores que emoções são expressas ali a partir de uma lista de seis a dez palavras para emoções na língua do observador. Como relatado por Ekman (1999b), experimentos desse tipo têm sido realizados com observadores de uma dúzia de países, revelando um importante alinhamento sobre quais emoções são exibidas (as taxas de reconhecimento são maiores para a felicidade, a tristeza e o nojo). Dito isso, a filiação a uma certa cultura aumenta o reconhecimento de expressões da cultura em questão, o que levou alguns a afirmarem que as expressões emocionais são uma linguagem universal com diferentes dialetos (ELFENBEIN *et al.*, 2007).

Somado a informações complementares sobre a produção de expressões faciais (MATSUMOTO *et al.*, 2008), os dados sobre reconhecimento têm sido considerados como favoráveis à hipótese de que os programas afetivos são adaptações evolutivas que geram as mesmas mudanças faciais, obrigatoriamente, em todas as culturas, embora regras de exibição específicas a cada cultura possam mascarar parcialmente essa universalidade transcultural.

As evidências para a universalidade têm sido criticadas com base em razões metodológicas e conceituais. Metodologicamente falando, tem-se argumentado que os experimentos são problemáticos na medida em que dependem de um paradigma de escolha forçado que superestima o consenso, e em que dependem de estímulos ecológicos irreais, tais como o rosto posado de atores (RUSSELL, 1994). Conceitualmente falando, argumenta-se que a hipótese evolutiva de que a seleção favoreceria a produção de certas expressões faciais obrigatórias é implausível, uma vez que em situações de conflito pode ser que não seja do interesse evolutivo que o sujeito emocionado permita a observadores saberem quais emoções ele está experimentando (por exemplo, comunicar medo durante um confronto). E mesmo que haja universalidade no reconhecimento e produção de expressões emocionais, explicações alternativas como a aprendizagem constante das espécies também poderiam dar conta das evidências citadas (FRIDLUND, 1994).

## **8.2 A teoria ecológico-comportamental, construtivismo psicológico e construtivismo social: emoções como construtos**

Uma alternativa influente à teoria das Emoções Básicas das expressões faciais é a teoria da Ecologia Comportamental (FRIDLUND, 1994), que substitui a

noção de *expressões de emoção* pela de *exibições* produzidas dependentemente do público, quando aqueles que as mostram esperam benefícios disso. A dependência de um público implica que aqueles que mostram as suas *exibições faciais* as moldam para o seu público, levando em conta o contexto, e não as produzem obrigatoriamente a partir da experiência de uma dada emoção. Tais *exibições* seriam, na verdade,

declarações que mostram a nossa trajetória em uma dada interação social, ou seja: o que faremos na situação presente, ou aquilo que gostaríamos que o outro fizesse (FRIDLUND, 1994, p. 130).

Por exemplo: o que Ekman descreve como sendo um rosto bravo, triste ou feliz é descrito pelos ecologistas comportamentais como a *exibição* de uma prontidão ao ataque, de uma intenção de aproximar-se e de um pedido de socorro, respectivamente (FRIDLUND, 1994). Vários teóricos afirmam que as abordagens de Ekman e Fridlund podem ser reconciliadas: as *emoções* podem ser, ao mesmo tempo, *expressões de emoções* e *declarações confiáveis* precisamente porque estão associadas com *emoções* (M. GREEN, 2007; SCARANTINO, 2017; HESS, BANSE; KAPPAS, 1995; BAR-ON, 2013).

O maior problema com a TBE tradicional é que os perfis de resposta distintiva supostamente produzidos em sequência não puderam ser convincentemente demonstrados nem no nível das respostas expressivas, nem no nível das mudanças autônomas, neurais, das ações pré-determinadas ou aprendidas (ORTONY; TURNER, 1990; MAUSS *et al.*, 2005; BARRETT, 2006; LINDQUIST *et al.*, 2012). Essa falta de assinaturas corporais claras e distintivas levou a várias tentativas de salvar a TBE de uma refutação empírica. Alguns teóricos das *emoções básicas* sugerem, por exemplo, que as *emoções básicas* podem ser reguladas, o que seria responsável por mascarar seu efeito obrigatório, ou que elas não podem ser provocadas num nível adequado apenas por estímulo laboratorial, ou que elas costumam misturar-se com outros estados cognitivos e afetivos, de modo que suas respostas distintivas acabam ficando confusas (EKMAN; CORDARDO, 2011).

Outros oferecem novas meta-análises mais favoráveis à existência de assinaturas biológicas específicas a *emoções*, sobretudo no nível das mudanças neurais e autônomas (KREIBIG, 2010; STEPHENS *et al.*, 2010; NUMMENMAA; SAARIMÄKI, no prelo). Uma terceira opção é fazer a transição para uma Nova Teoria Básica das *Emoções*, que substituísse a suposição das respostas em



sequência pela de tendências para a ação com precedência de controle, o que poderia explicar algo da variabilidade das respostas, ao mesmo tempo que preservaria a ideia básica de que as emoções são orientadas para resolver problemas evolutivos (LEVENSON, 2011; SCARANTINO; GRIFFITHS, 2011; SCARANTINO, 2015).

Uma proposta mais radical foi oferecida pelos construtivistas psicológicos, que sugerem que nós deveríamos abandonar completamente o modelo de "variável latente" distintivo da teoria básica das emoções, substituindo-o por um modelo de "variável emergente" segundo o qual as emoções não causariam expressões faciais, mudanças autônomas e ações pré-determinadas e adquiridas, mas emergiriam delas (BARRETT; RUSSELL, 2015). Especificamente, os construtivistas psicológicos afirmam que não há uma correspondência unívoca entre raiva, medo, felicidade, tristeza etc., e quaisquer respostas neurobiológicas, fisiológicas, expressivas, comportamentais ou fenomenológicas, e que as diferentes respostas supostamente indicativas das emoções básicas não estão sequer fortemente correlacionadas a estas.

Construtivistas psicológicos concluem que essa variabilidade põe em xeque a própria ideia de que

as emoções tenham um estatuto ontológico de entidades causais [e que] elas existam no cérebro ou no corpo, causando mudanças *nas saídas* sensória, perceptiva, motora ou fisiológica (BARRETT, 2005, p. 257).

Essa ideia encontra-se em direta oposição à Teoria Motivacional, segundo a qual as emoções são motivos – determinantes causais das mudanças de *saídas* que observamos.

Também já foi dito que as categorias da psicologia popular [*folk psychology*] normalmente invocadas pelos teóricos das emoções básicas – por exemplo: raiva, medo, nojo, etc. – não são objetos próprios à investigação científica, e que deveriam ser substituídas por categorias que descrevam os componentes emocionais, em vez das distintas emoções elas mesmas (RUSSELL, 2003; BARRETT, 2006, 2017).

Os construtivistas estão convencidos de que as emoções surgem com o andar da carruagem, de modo flexível, usando blocos de construção que não são específicos a emoções – mais ou menos como nossas refeições são construídas a partir de ingredientes que não são específicos a elas, e que poderiam ser usados de acordo com diferentes receitas. Um dos ingredientes com os quais as emoções seriam feitas é o *afeto-central*, que é

um estado neurofisiológico conscientemente acessível como uma sensação simples, não-reflexiva; uma mistura completa de valores hedônicos (prazer-desprazer) e despertadores (dormente-ativado). (RUSSELL, 2003, p. 147)

Os construtivistas psicológicos enfatizam que sempre estamos em algum estado de afeto-central, o que seria um tipo de barômetro a nos informar sobre nossa "relação" com o fluxo de eventos. As leituras desse barômetro são sentimentos, entendidos como uma mescla de prazer-desprazer e ativação-desativação. Tais leituras poderiam ser representadas como pontos ao longo de uma "estrutura circunflexa", em que o eixo vertical representa o grau de ativação-desativação, e o eixo horizontal representa o grau de prazer-desprazer (RUSSELL, 1980):



Diferentes construtivistas descrevem a maneira pela qual as emoções são construídas a partir do afeto-central e outros ingredientes, de jeitos diferentes. Por exemplo: na influente Teoria do Ato Conceitual de Barrett, a conceituação desempenha um papel-chave (BARRETT, 2006, 2013, 2017; BARRETT; SATPUTE, 2013). Ter medo significaria categorizar um estado afetivo cerne de alta ativação e desprazer

sob o conceito de "medo". Estar feliz significaria categorizar um estado afetivo cerne de alta ativação e prazer sob o conceito de "felicidade". Mais genericamente, Barrett (2017) considera que as emoções são experiências que emergem da categorização das sensações do próprio corpo e do mundo. Essa abordagem, que lembra a teoria da cognição-excitação de Schachter e Singer (1962) e acrescenta temas das tradições sentimentalista e avaliativa, tem sido criticada por confundir as emoções com rótulos verbais, tornando impossível a humanos adultos rotular equivocadamente suas próprias emoções, e por impedir crianças e animais de terem emoções, sobretudo (*vide* SCHERER, 2009). Confira Barret (2015) para uma resposta.

Russell (2003) considera que a conceituação afeta apenas a **meta-experiência** da emoção, por exemplo: a tomada de consciência de que se tem medo, e reconhece que os episódios emocionais possam ser construídos sem o envolvimento de categorização. Para ele, há uma variedade de mecanismos causais independentes, em vez de um mecanismo qualquer, específico às emoções, que explicam por que há um certo grau de correlação entre mudanças expressivas, autônomas e comportamentais em episódios emocionais, ainda que ele enfatize que as correlações são bastante mais fracas do que o modelo de Ekman (1999a) poderia prever (RUSSELL, 2012).

Recentemente, algumas propostas foram apresentadas a fim de integrar o construtivismo psicológico com outros programas de pesquisa. Alguns sugerem que o progresso estará em combinar a teoria da apreciação com a versão do construtivismo psicológico de Russell, tendo apresentado uma teoria geral de como as tendências para a ação emocional são causadas pela ponderação das expectativas de utilidade das opções de ação (MOORS, 2017). Outros propõem que nós distingamos claramente o lado motivacional e fenomenológico dos fenômenos afetivos, deixando o lado motivacional de (algumas) emoções para uma nova teoria dos circuitos de sobrevivência, e preservando os termos da psicologia popular das emoções exclusivamente para designar sentimentos, entendendo os últimos como cognitivamente construídos (LEDOUX, 2015, 2017). Note a diferença em seu trabalho anterior (LEDOUX, 1996).

Outra opção que possui elementos em comum com o construtivismo psicológico é o construtivismo social. A abordagem dos construtivistas sociais teve seus primeiros defensores nos anos 1920, quando muitos antropólogos e cientistas sociais começaram a questionar as evidências de Darwin (1872) para a universalidade das expressões emocionais (ALLPORT, 1924; LANDIS, 1924; KLINEBERG, 1940).

Esses pesquisadores iniciaram o ramo do construtivismo social que podemos chamar de "variabilidade cultural", cuja ideia é a de que as emoções são diferentes em vários aspectos essenciais em diferentes culturas. Essas diferenças têm sido apontadas tanto em relação ao léxico das emoções (RUSSELL, 1991; WIERZBICKA, 1999) como também em relação às características indicativas das emoções (MESQUITA; FRIJDA, 1992; MESQUITA; PARKINSON, no prelo).

A corrente do construtivismo social mais semelhante à tradição motivacional é a corrente do "papel social", cuja tese é a de que as emoções cumprem funções sociais razão pela qual elas devem ser consideradas ações ou papéis, em vez de paixões. Sobre o "mito das paixões", confira, por exemplo, Solomon (1976) e Averill (1990). Jean Paul Sartre (1939 [1948]) pode ser considerado como o primeiro a ter apresentado uma teoria geral, ainda que idiossincrática, das emoções como papéis sociais, uma ideia desenvolvida nos anos 1980 pelos filósofos (HARRÉ, 1986; ARMON-JONES, 1986), psicológicos (AVERILL, 1980) e antropólogos (LUTZ, 1988). Recentemente, Parkinson (1995, 2008, 2009), Parkinson, Fischer e Manstead (2005), Griffiths (2004), Mesquita e Boiger (2014) e Van Kleef (2016) articularam sofisticadas explicações de tipo construtivista social incorporando a essa tradição temas das teorias evolutivas.

## 9. A teoria motivacional na filosofia recente

Há duas versões da Tradição Motivacional na filosofia contemporânea das emoções. A *versão fenomenológica*, articulada por Deonna e Teroni (2012, 2015), considera que as emoções são sensações de prontidão para a ação. A *versão não-fenomenológica*, articulada por Scarantino (2014, 2015), identifica as emoções como causas de estados de prontidão para ação que podem ou não serem sentidos. Ambas as versões concordam que o aspecto fundamental de uma ação é o modo pelo qual ela motiva o emocionado a agir.

Deonna e Teroni defendem que tanto a teoria julgamentalista quanto as teorias perceptivas da emoção cometem o erro de caracterizar as emoções em termos de **conteúdo**, e não de **atitude** ou **modo**. Como Searle (1979, p. 48) aponta, "todos os estados intencionais consistem em um conteúdo representativo em um modo psicológico". Por exemplo: acreditar e desejar são tipos diferentes de modos ou atitudes psicológicos, e cada um deles tem um conteúdo – aquilo que é acreditado ou desejado enquanto capturado por uma proposição, respectivamente.

Se as emoções fossem tipos especiais de juízos ou percepções, elas seriam diferentes de outros tipos de juízos ou percepções não em termos da atitude, mas apenas em termos do conteúdo – o que é julgado ou percebido quando nos emocionamos. Além disso, as próprias emoções seriam diferentes umas das outras apenas em termos do conteúdo, em vez da atitude, já que não haveria atitude específica para, digamos, a raiva, a vergonha, a culpa e assim por diante, mas apenas para uma atitude comum – a atitude judicativa ou perceptiva – com relação a diferentes conteúdos. Deonna e Teroni (2015) acreditam que essa ideia fracassa em capturar não apenas o que diferencia as emoções umas das outras, mas também o que as torna especiais enquanto estados motivacionais.

Como alternativa, eles propõem uma *teoria atitudinal das emoções*. Segundo ela, o medo de um tigre não é nem o julgamento, nem a percepção de que há algo perigoso por perto, mas a atitude de tomar-como-perigoso direcionada ao conteúdo de que há um tigre. O que daria conteúdo às atitudes emocionais, dizem Deonna e Teroni, são suas **bases cognitivas**, os modos pelos quais fazemos a cognição do conteúdo de que há um tigre – digamos, pela percepção, imaginação, inferência e assim por diante (por exemplo, a percepção de que há um tigre).

Mas que tipo de atitude é essa que constitui uma emoção em vez de, digamos, um juízo ou uma percepção? Deonna e Teroni consideram que as atitudes emocionais sejam experiências em que sentimos o nosso corpo como pronto para a ação. Por exemplo: o medo de cachorro equivalerá à "experiência do cachorro como perigoso" na medida em que seja "uma experiência do próprio corpo como preparado" para evitar e afastar-se (DEONNA; TERONI, 2015, p. 303). Do mesmo modo, ter raiva de alguém "é uma experiência de ofensa na medida em que consiste na experiência do próprio corpo como preparado para retaliar" (2015, p. 303). Assim, ter emoções é sentir atitudes de prontidão para a ação que são irreduzíveis a atitudes não-emocionais e específicas a cada emoção. Para uma crítica à teoria atitudinal, confira Rossi e Tappolet, no prelo.

O ponto de partida para a *Teoria Motivacional das Emoções* de Scarantino (2014) é a convicção de que as emoções são irreduzíveis não só a juízos e percepções, mas também a sensações, e que na verdade deveriam ser entendidas como tipos especiais de **estados centrais de motivação** ou **programas comportamentais**. Estados centrais de motivação ou programas comportamentais são definidos pelo que fazem, e não por como os sentimos. E o que eles fazem é fornecer "uma orientação geral para o comportamento ao potencializar seletivamente conjuntos coerentes de opções de comportamento" (GALLISTEL, 1980, p. 322).

A potenciação seletiva pode resultar em sensações, mas mudanças fenomenológicas não são necessárias para a própria potenciação, que consistiria em mudanças nas probabilidades das opções de comportamento. Um exemplo é o medo, que envolveria a potenciação seletiva de opções compartilhando o objetivo de evitar um certo alvo, apreciado como perigoso. A raiva envolveria a potenciação seletiva das opções que compartilham o objetivo de atacar um certo alvo apreciado como ofensivo. A culpa envolve a potenciação seletiva das opções que compartilham o objetivo de consertar um relacionamento apreciado como danificado por ações aquém de nossos padrões morais, e assim por diante.

A teoria motivacional das emoções inspira-se na teoria das emoções como tendências a agir de Frijda (1986), mas há diferenças. Scarantino (2014, 2015) estabelece uma distinção entre uma **emoção** e um **episódio de emoção**, em que a primeira corresponde ao que causa uma mudança na prontidão para a ação e o episódio de emoção corresponde à mudança de fato na prontidão para a ação. Mas Scarantino bebe da fonte da teoria de Frijda (1986) sobretudo quanto à suposição de que as tendências para a ação devem ter uma **precedência de controle** para serem emocionais. Precedência de controle envolve interromper processos concorrentes, garantir acesso – na memória, inferência, percepção etc. – às informações não relacionadas ao objetivo de evitar/afastar-se e preparar o corpo para a ação.

A ideia de que as emoções são programas comportamentais que causam impulsos prioritários para se agir pode ser combinada com uma **história da origem** de como algumas dessas programações evoluíram de modo a lidar com as tarefas fundamentais da vida, levando ao que Scarantino (2015) chama de Nova Teoria Básica da Emoção. Segundo esta teoria, a aprendizagem pode afetar tanto aquilo que ativa o programa evoluído (**entrada**) quanto as respostas que o programa causa (saída), graças à interação entre a regulação e a tendência à ação prioritária. Isso resultaria em uma enorme variabilidade de respostas reais observadas na ativação de qualquer emoção básica, protegendo a nova TBE do problema da falta de “assinaturas corporais”.

Por fim, Scarantino (2014) endossa a teoria teleossemântica do conteúdo a fim de que as emoções possam lidar com o problema da intencionalidade, propondo que diferentes emoções se diferem umas das outras e também de estados não-emocionais **tanto** em termos do estado de tendência à ação prioritária que elas causam (a atitude) quanto em termos do que elas representam (o conteúdo). Nessa abordagem, o medo é uma programação de controle de ação prioritária que representaria perigos porque tem a função de causar comportamentos de evasão

na presença de um perigo; a raiva é um programa de controle de ação prioritária que representaria desprezos porque tem a função de causar comportamentos de agressividade na presença do desprezo, e assim por diante.

Um desafio central para as correntes fenomenológica e não-fenomenológica das teorias motivacionais das emoções é explicar os estados de prontidão para a ação característicos de diferentes emoções. Primeiramente, muitas emoções não parecem motivar quaisquer ações. Luto e depressão, por exemplo, parecem envolver um grande enfraquecimento da prontidão para a ação. Em segundo lugar, não fica claro que tendências para a ação as emoções “retrospectivas” (como o arrependimento) poderiam provocar, já que focam apenas naquilo que aconteceu no passado, e que não poderia ser mudado. Em terceiro lugar, emoções como a alegria envolvem potenciação seletiva de um leque bastante grande de opções comportamentais, no que também não fica claro quais tendências para a ação deveriam ser associadas a elas. Em quarto lugar, parece possível que uma mesma tendência para a ação seja associada a diferentes emoções, e que diferentes emoções sejam associadas à mesma tendência para a ação, se tais tendências são descritas em um nível de análise suficientemente abstrato. Para críticas às tendências motivacionais confira, por exemplo, Reizenstein (1996), Prinz (2004), Tappolet (2010, 2016) e Eder e Rothermund (2013).

## 9.2 Teorias enativistas das emoções: emoções como "enações"<sup>40</sup>

O enativismo é um programa de pesquisa interdisciplinar que começa com uma insatisfação com a forma pela qual os processos cognitivos têm sido estudados nas ciências cognitivas (DI PAOLO; THOMPSON, 2014; GALLAGHER, 2017). Dois temas enativistas, em particular, são importantes para uma teoria das emoções. O primeiro é o foco no papel ativo desempenhado pelo agente de cognição em sua relação com o mundo exterior, que para os enativistas não é pré-dado, nem passivamente detectado, mas é na verdade **atuado** [*enacted*] e ativamente moldado pelos poderes de “*sense-making*”<sup>41</sup> do agente cognitivo. Essa atividade de *sense-*

---

<sup>40</sup> N.T.: *Enactions*, no original em inglês. O termo é um neologismo recente, mas na falta de melhor tradução, também poderíamos falar em “atuação” (no sentido de ação ativa).

<sup>41</sup> N.T.: Termo técnico enativista ainda sem tradução.

*making* é o coração da cognição tal qual os enativistas a entendem, e é feita por todos os seres vivos, por mais simples que sejam, na medida em que são sistemas autônomos e adaptativos (THOMPSON, 2007).

O segundo tema é o foco no caráter **incorporado, situado e estendido** dos processos cognitivos. O tema da corporeidade ronda boa parte da ciência dos afetos também. A esse respeito, confira, por exemplo, Niedenthal (2007), Wilson-Mendenhall *et al.* (2011) e Carr *et al.*, no prelo. Enquanto as ciências cognitivas tradicionais e a neurociência têm focado no cérebro, em separado do resto do corpo e do ambiente, os enativistas afirmam que fracassaremos em entender a cognição se negligenciarmos as interações causais recíprocas entre o cérebro, o corpo e o ambiente que ocorrem dinamicamente ao longo do tempo.

A ideia de que habilidades cognitivas complexas podem depender de um cimbramento garantido pelo ambiente externo tem sido especialmente popular entre os teóricos da emoção. Ela já levou, por um lado, a uma renovada atenção ao papel desempenhado pela comunicação interpessoal em ambientes sociais (GRIFFITHS; SCARANTINO, 2009), e, por outro, à sugestão de que as emoções são ontologicamente estendidas além dos limites estreitos do crânio (STEPHAN *et al.*, 2014; KRUEGER, 2014; COLOMBETTI; ROBERTS, 2015; COLOMBETTI, 2017).

A qual tradição de pesquisa os enativistas pertencem? O foco na experiência parece empurrar o enativismo na direção da Tradição Sentimentalista (ver, por exemplo: RATCLIFFE, 2008). Com efeito, o enativismo é de fato influenciado pela ideia, central à tradição filosófica fenomenológica, de que o corpo é uma estrutura experienciada (HUSSERL, 1952 [1989]; MERLEAU-PONTY, 1945 [1962]) em vez de uma estrutura **física**, apenas. E o que podemos experimentar limita o mundo em que habitamos, o nosso "*Umwelt*" (UEXKÜLL, 1934 [2010]).

Colombetti (2014) afirma que a tradição fenomenológica pode enriquecer a neurociência das emoções. Apoiando-se no trabalho de *neurofenomenologia* de Varela (1996), Colombetti desenvolveu uma matriz que integra métodos em terceira pessoa, como o escaneamento cerebral, com métodos em primeira pessoa, como autorrelatos. Também fica bastante claro que os enativistas, em contraste com a "posição desencarnada" (COLOMBETTI; THOMPSON, 2008) de muitas teorias cognitivistas, encaram as emoções como processos corporais e experienciais, e não intelectuais.

Ainda assim, é mais apropriado encaixar o movimento enativista no lado fenomenológico da Tradição Motivacional. Isso se dá porque os enativistas também enfatizam fortemente o papel da **ação** na cognição. Muitos deles têm apresentado explicações das emoções enfatizando sua natureza orientada para a ação



(HUFUNDIEK, 2016; SLABY; WÜSCHNER, 2014; SHARGEL; PRINZ, 2018). A cognição é dita ser **atuada** por sistemas vivos inerentemente teleológicos para fins de ação. Mais radicalmente, alguns processos cognitivos, como a percepção, são inclusive descritos como constitutivamente dependentes de processos motores, como na teoria sensório-motora da consciência visual (HURLEY, 1998; O'REGAN; NOË, 2001).

Não há entre os enativistas um entendimento uno da relação entre as emoções e a ação, mas um grande número de propostas diferentes. Na obra de Colombetti, por exemplo, a noção de *auto-organização* desempenha um papel central. Segundo ela, episódios emocionais são "padrões de auto-organização do organismo, melhor descritos com as ferramentas conceituais da teoria dos sistemas dinâmicos" – um ramo da matemática destinado a explicar a evolução temporal de sistemas que mudam ao longo do tempo (COLOMBETTI, 2014, p. 53; *vide* LEWIS, 2005).

A auto-organização seria a capacidade de um sistema complexo em alcançar e manter um estado de ordem através de influências causais recíprocas entre partes componentes mais simples. Quando aplicada às emoções, a ideia é que os componentes das emoções auto-organizam-se, o que ajudaria a explicar a variabilidade dos episódios emocionais, já que sistemas auto-organizadores podem ter múltiplos estados-fim, dependendo de como seus componentes interagem (*vide* A. CLARK, 2001, p. 113-114).

Ainda que haja analogias entre essa teoria e o construtivismo psicológico, especialmente no que diz respeito à ênfase nas emoções como fenômenos flexíveis e emergentes, Colombetti nega que atos conceituais façam surgir emoções, afirmando que a "produção de sentido" [*sense-making*] é um fenômeno muito mais primitivo, presente já em bactérias e chegando aos seres humanos. Criaturas envolvem-se em "produção de sentido" quando avaliam o seu ambiente em termos de se ele promove a sua automanutenção, agindo assim de modo a melhorar a sua viabilidade dentro desse ambiente, como uma bactéria faz quando nada para longe de uma substância nociva.

Ao mesmo tempo, Colombetti usa a suposição da auto-organização de fenômenos emocionais para se opor à ideia de que episódios emocionais são causados por programas afetivos (contra a tradição das emoções básicas) ou por apreciações (contra a tradição da apreciação). Incidentalmente, Colombetti (2014) acredita que a própria noção de emoção básica é arbitrária e indigna de ser mantida, pois desencoraja a pesquisa das características neural, comportamental e corporal das emoções supostamente não-básicas. Hufendiek (2016) apresenta a ideia

complementar de que emoções supostamente não-básicas mostram muitas características distintivas das emoções básicas (ver também J. CLARK, 2010).

Outro ponto distintivo do enativismo é sua posição *antirrepresentacional* (VARELA *et al.*, 1991; HUTTO; MYIN, 2013; GALLAGHER, 2017). Hutto (2012), por exemplo, propõe que "abandonemos a ideia de que as emoções representam situações em maneiras veridicamente avaliáveis" (2012, p. 4), sugerindo que as emoções não representam temas relacionais centrais. O medo, por exemplo, não representaria que há um perigo próximo, e a raiva não representaria que houve um desprezo contra mim ou os meus próximos. A principal preocupação de Hutto (2012) com a atribuição de poderes representacionais às emoções é que tais poderes são postulados apesar de não terem nenhuma força explicativa (ver HUFENDIEK, 2018, para uma discussão). Mais especificamente, Hutto (2012) segue os passos de Ramsey (2007) em assumir que um estado mental conta como representacional apenas se for consumido por outros sistemas à luz daquilo que diz ou indica, concluindo que as emoções falham em desempenhar esse papel explicativo na economia cognitiva do organismo, não devendo serem consideradas como representações, portanto.

Prinz (2004) costumava acreditar que as emoções representassem temas relacionais centrais porque tinham a função de correlacionarem-se com eles, mas ele mudou de ideia em seus trabalhos mais recentes. Schargel e Prinz (2018) defendem que a abordagem teleossemântica é uma ameaça ao caráter verdadeiramente incorporado de uma teoria das emoções à moda James-Lange – uma abordagem preferida por eles. Isso se dá porque qualquer veículo não-incorporado, por exemplo, um juízo desencarnado, que tem a função de correlacionar-se a um tema relacional central representaria tal tema tão bem quanto um veículo incorporado o faz (SHARGEL, 2014).

Como alternativa, Shargel e Prinz (2018) adotam um teoria enativista não-representacional do conteúdo das emoções segundo a qual as emoções criam possibilidades de ação graças à preparação corpórea que elas envolvem (*vide* GRIFFITHS; SCARANTINO, 2009; HUFENDIEK, 2016). Tais possibilidades de ação, diferentemente das *affordances* da tradição gibsoniana, que preexistem às emoções e são motivacionalmente inertes, são "dependentes de estados (elas tipicamente surgem apenas depois que a emoção já começou), e imperativas (elas motivam a ação)" (SHARGEL; PRINZ, 2018, p. 119). Segundo esse ponto de vista, o medo gera possibilidades para a fuga que não se dariam na ausência de medo, e que funcionam como atratores dinâmicos, atraindo o agente em direção à fuga. O conteúdo atuado do medo, portanto, não é o perigo, mas a apresentação de uma

certa situação como algo de que se deve escapar, junto de um impulso de mover-se para longe – um conteúdo que é essencialmente incorporado, já que envolve uma preparação corporal para a fuga.

Um desafio central para as teorias enativistas das emoções de corrente não-representacional é explicar nossas práticas normativas com respeito às emoções. Uma vez que nos damos conta de que o medo de alguém levou ele ou ela a evitar um certo estado de coisas, ou que a raiva de alguém motivou ele ou ela a atacar alguém, ainda nos perguntamos se o que motivou a evasão é um perigo ou não, e se o que motivou a retribuição é um desprezo ou não. Em outras palavras, ainda tratamos as emoções como adequadas ou inadequadas relativamente às circunstâncias de sua ocorrência. É uma questão em aberto se e como essas formas de adequabilidade podem ser compreendidas se as emoções não representam temas relacionais centrais. Para um aprofundamento, confira Hufendiek (2016, 2017, 2018).

## 10. Racionalidade e emoções

### 10.1 Racionalidade cognitiva como adequação, justificação e coerência

Distinguímos a *racionalidade cognitiva* das emoções, que consiste na habilidade das emoções de representar como o mundo é e relacionar-se devidamente a outros processos avaliativos sensíveis às evidências, da *racionalidade estratégica* das emoções, que consiste na habilidade das emoções de guiar as ações que promovem os interesses do agente e relaciona-se devidamente a outros processos influentes sobre as ações (DE SOUSA, 1987, 2011; GREENSPAN, 2000; MULLIGAN, 1998; SOLOMON, 1980; THAGARD, 2006; STEPHAN, 2017B).

As emoções têm sido há muito tempo mal-entendidas tanto no que diz respeito à sua racionalidade cognitiva quanto estratégica. Os estoicos são famosos por terem dito que as emoções são falsos juízos. Por exemplo: ter medo de um tigre envolveria o falso juízo de que a própria vida em perigo seria algo importante, enquanto o sábio deveria ser indiferente a isso e a tudo o mais – menos à virtude. O fracasso das emoções em um nível estratégico também é encontrado profundamente arraigado tanto no senso comum quanto nas abordagens teóricas. *Ira brevis furor*,

diziam os romanos<sup>42</sup>: a raiva é um breve acesso de loucura. Em tempos recentes, o pêndulo moveu-se para o outro lado, e pesquisadores da filosofia e da ciência dos afetos têm começado a reabilitar as emoções nos termos de sua racionalidade tanto cognitiva quanto estratégica. Uma consideração apropriada do papel das emoções no que diz respeito à sua racionalidade exige algumas distinções.

A primeira dessas distinções diz respeito às três variedades de racionalidade cognitiva das emoções: *racionalidade como adequação*, *racionalidade como justificação* e *racionalidade como coerência*. A tese dominante sobre as emoções é que elas são representações de temas relacionais centrais ou objetos formais. Logo, a primeira dimensão da avaliação dessa racionalidade diz respeito a se tais temas relacionais centrais/objetos formais são ou não instanciados. Podemos dizer, por exemplo, que o medo é **racional em termos de adequação** caso ele seja direcionado a coisas que realmente sejam perigosas, pois é isso que o medo representa. Temer um tubarão nadando à nossa volta é algo adequado, já que tubarões são perigosos. Como D'Arm e Jacobson (2000) enfatizam, é importante notar que essa adequação se difere de formas de adequação que são morais ou estratégicas. Por exemplo: divertir-se com uma piada pode ser adequado ainda que divertir-se por isso seja moralmente inadequado devido a um conteúdo sexista da piada, ou custoso em termos do interesse próprio, já que aqueles que testemunharem o regozijo podem vir a formar uma má impressão do agente que se diverte.

Suponha agora que algo que não é perigoso provoque medo. O medo ainda assim poderia manifestar uma *racionalidade como justificação* se seu objeto particular exibir importantes pistas de sua periculosidade. Por exemplo: temer uma réplica realista de um tubarão movendo-se à nossa volta dentro da água seria racional no sentido de estar justificado, ainda que, sem sabermos, a réplica do tubarão fosse controlada remotamente por um grupo de inofensivos biólogos marinhos.

Uma terceira dimensão da racionalidade cognitiva diz respeito à **consistência** das emoções com outras representações de como o mundo é. Se alguém tem a experiência de ter medo de voar e acredita que voar é perigoso, há uma racionalidade aí, entendida como coerência entre o que é temido e o que se acredita (ainda que o medo seja inadequado dada a probabilidade extremamente baixa de acidentes aéreos).

Como observado na seção 7.1, as emoções normalmente são **resistentes a razões**: muitas pessoas **não** acreditam que voar seja perigoso, e mesmo assim continuam a ter medo disso. Mas as emoções manifestam racionalidade enquanto coerência em uma variedade de situações. E isso é o caso porque o fato de que

---

<sup>42</sup> N.T.: Máxima de Horácio encontrada nas *Epístolas* (1, 2, 62).

as emoções têm **bases cognitivas**, que consistem em cognições cuja função é fornecer às emoções seus objetos particulares – eu devo crer, perceber ou imaginar estar em um avião antes de ter medo disso.

Quando tais cognições são crenças, a sua modificação tende a ser coerentemente refletida em mudanças nas emoções. Por exemplo: se eu estou bravo com Tom com base na minha crença de que ele falou mal de mim para um colega, minha raiva provavelmente não persistirá se e quando eu me der conta de que, na verdade, ele não falou mal de mim para um colega. Pode ser, no entanto, que minha raiva com Tom persista apesar de minha crença que não houve menosprezo algum, revelando assim minha irracionalidade como coerência.

Um caso especial de racionalidade como coerência diz respeito à coerência dos conjuntos de emoções. Helm (2009) afirma que as emoções vem em padrões racionais articulados em torno das coisas que importam para o agente. Por exemplo: se evitar a morte é uma preocupação minha, então eu deveria não só ter medo de quando minha vida fosse ameaçada por uma doença mortal, mas também deveria sentir alívio, sob pena de irracionalidade, uma vez afastada essa ameaça, e tristeza ou desilusão se a doença viesse a progredir.

## 10.2 Racionalidade estratégica instrumental e substantiva

A *racionalidade estratégica (ou prudencial)* das emoções diz respeito à habilidade delas de orientar ações que promovam os interesses do agente e de relacionar-se devidamente a outros processos que influem nas ações, em especial na tomada de decisão. Ainda que as emoções que são estrategicamente racionais em geral também sejam cognitivamente racionais no sentido da adequação e da justificação, há exceções. Por exemplo: alguns casos de raiva que não ocorrem na presença de menosprezos reais (adequação) ou de indícios de menosprezo (justificação) podem acabar promovendo os interesses do agente. Um exemplo disso seria a raiva de um cliente cujos interesses seriam melhor atendidos se pudesse devolver a mercadoria, mas que, tendo perdido a nota fiscal, acaba dando uma bronca raivosa no vendedor, que acaba cedendo por ter ficado intimidado.

Podemos distinguir dois componentes da racionalidade estratégica: uma emoção é racionalmente estratégica desde que leve um agente a (i) selecionar meios propícios aos fins do agente (racionalidade estratégica instrumental) e (iii) buscar fins que se alinhem com os interesses do agente, considerado todo o resto

(racionalidade estratégica substantiva). Um exemplo de irracionalidade instrumental seria a de um agente que entra em pânico enquanto tenta sair de uma casa em chamas, não consegue escutar as orientações do bombeiro, vai para a porta mais próxima, esquecendo que ela leva à única área da casa sem uma saída, e morre no processo. Não há nada de errado aqui com o fim a que o pânico predispõe o agente – encontrar segurança – mas os meios escolhidos são claramente menos do que ideais.

Pode-se argumentar que a irracionalidade substantiva encontra-se tanto nos tipos quanto **nas ocorrências** de emoções. Por exemplo: alguns afirmam que o luto é um **tipo** de emoção substantivamente irracional, pois sempre envolve a crença de que a pessoa querida ao enlutado está morta, bem como o desejo de que tal pessoa não esteja morta, o que é um desejo insatisfazível, considerando o que a pessoa acredita (GUSTAFSON, 1989). Confira Cholbi (2017) para uma resposta.

Não mais das vezes, o que os teóricos afirmam é que ocorrências específicas de certas emoções podem ser substantivamente irracionais. Por exemplo: seria contrário aos interesses de um agente que ele ficasse irado com um potencial empregador durante uma entrevista de emprego, já que isso provavelmente resultaria em não conquistar a vaga e assim frustrar seu interesse próprio.

Por outro lado, ter raiva de alguém que fura a fila em um aeroporto pode ser substantivamente racional, já que o objetivo de interromper um comportamento ofensivo serve aos nossos interesses. Nussbaum (2016) defende que isso só pode ser o caso se o foco do agente irado não for, em caso algum, o de retribuir o comportamento ofensivo, mas o de prevenir que o comportamento ofensivo aconteça novamente, já que o desejo de retribuição é claramente irracional ou problemático de diversas maneiras (Nussbaum refere-se às formas não-problemáticas de raiva como *raiva de transição*).

Uma razão comum para duvidar das estratégias de racionalidade das emoções é que elas normalmente parecem levar a ações físicas e mentais **impulsivas**. A impulsividade envolve agir rapidamente antes de considerar todas as informações relevantes (FRIJDA, 2010; ELSTER, 1999, 2010). Alguns afirmam que isso é precisamente o que possibilita que as emoções sejam uma solução ótima entre velocidade e flexibilidade, permitindo que as emoções funcionem como “reflexos desacoplados” (SCHERER, 1984).

Outros observam que as emoções costumam levar a ações “não-rationais [*arational*]”, especialmente ações emocionais não executadas “com alguma razão” (HURSTHOUSE, 1991). Exemplos paradigmáticos incluem ações como pular

sem parar por alegria, ou rolar sobre as roupas da esposa morta por luto. Em casos assim, diz Hursthouse, não há par de crença e desejo que possa ser apresentado a fim de fornecer uma razão humana para tais ações, que deveriam ser explicadas apenas dizendo que o agente está sob controle de uma emoção. O debate sobre ações não-rationais está aceso na filosofia recente das emoções, e uma variedade de propostas, tanto a favor como contra o humanismo, estão disponíveis (M. SMITH, 1998; GOLDIE, 2000; DÖRING, 2007; KOVACH; DE LANCEY, 2005; SCARANTINO; NIELSEN, 2015).

As emoções são notórias por nos fazerem agir de jeitos que acabamos por lamentar. Notavelmente, elas podem ser uma fonte de fraqueza da vontade, um fracasso em se agir de acordo com as próprias melhores razões (DAVIDSON, 1970 [1980]). Mas elas também podem igualmente ajudar um agente a persistir em seus objetivos de longo prazo (TAPPOLET, 2016, p. 227). Por exemplo: um forte sentimento de culpa pode ajudar um agente a resistir à tentação de trair o cônjuge, e, nesse sentido, ser um meio para o fim de exercer com sucesso o autocontrole. De vez em quando as emoções podem inclusive servir de base para o fenômeno da acrasia<sup>43</sup> inversa (MCINTYRE, 1990; ARPALY; SCHRODER, 2000), que consiste em não fazer o que se julga como o melhor, apenas para descobrir que na verdade se fez o que era melhor para si, diferentemente do que se havia pensado anteriormente. Por exemplo: podemos achar que é melhor nos tornarmos um músico profissional, mas ficarmos alijados pelo medo de palco e assim acabarmos indo para a faculdade de direito – e descobrir só mais tarde que esse curso de ação foi o que melhor serviu aos nossos interesses de longo prazo.

Uma outra ameaça à estratégia da racionalidade das emoções vem da sua relação com o autoengano (FINGARETTE, 1969; MELE, 1987; VAN LEEUWEEN, 2007), que é normalmente visto como irracional. Em resumo, o autoengano envolve a formação de crenças que são contrárias ao que indica a evidência à nossa disposição, mas conforme ao que o agente que se autoengana deseja. Emoções podem causar autoengano porque podem levar a poderosos desejos de que algo seja ou não seja o caso, o que por sua vez pode impactar causalmente a habilidade do sujeito de processar evidências.

---

<sup>43</sup> N.T.: Do grego *akrasía*, o estado de espírito em que se age contra o que se julga como o melhor, por causa de uma fraqueza da vontade.

Essa característica está sobretudo ligada ao fato de que as emoções determinam a saliência de possíveis objetos de atenção. Os poetas sempre souberam que o principal efeito do amor é redirecionar a atenção: quando amo, nada noto exceto o ser amado – e nenhum de seus defeitos. Mas isso traz um risco, pois posso não perceber que há enormes evidências de que estou sendo enganado de uma maneira prejudicial. Meu desejo de não ser enganado, motivado por meu amor, é o que causa esse processamento defeituoso das evidências, levando ao autoengano.

Esse potencial que as emoções têm de “distorcer a paisagem epistêmica” (GOLDIE, 2004, p. 259) negativamente é compensado pelo importante papel das emoções em promover o pensamento epistêmico racional. *Emoções epistêmicas* são aquelas particularmente relevantes para a nossa busca por conhecimento e entendimento. A curiosidade motiva a investigação, o interesse nos mantém nela, e, como Platão e Descartes observaram, a dúvida é crucial para a nossa habilidade de evitar o preconceito. Tais emoções “epistêmicas” podem nos guiar particularmente nas situações em que tentamos obter conhecimento (SILVIA, 2006; BRUN, DOĞUOĞLU; KUENZLE, 2008; MORTON, 2010).

Mas mesmo emoções banais, não epistêmicas, podem promover entendimento do mundo e de si mesmo. Confirma, também, a hipótese do sentimento-como-informação na ciência dos afetos, proposta por Schwarz (2012). De acordo com Brady (2013), a principal maneira pela qual as emoções fazem isso é motivando-nos a buscar informações relevantes sobre a adequação de nossas emoções, e sobre a adequação de seus interesses subjacentes. Uma vez mais, o dispositivo é o de mudar a saliência dentre possíveis objetos de atenção. Suponha, por exemplo, que você sinta medo logo antes de ter que propor um brinde em um casamento. O seu medo promove entendimento porque lhe provoca a ver se a situação é realmente perigosa ou se você está preocupado demais em dar um discurso brilhante.

De modo mais geral, alguns afirmam que a habilidade das emoções em mudar a atenção para alguma coisa em vez de outra é uma solução essencial para o chamado *problema do enquadramento* [*frame problem*], que é o desafio de separar informações relevantes de informações irrelevantes durante uma tomada de decisão. De Sousa, na filosofia, defende que

as emoções nos poupam de uma potencial paralisia provocada pelo [problema do enquadramento], controlando a saliência das coisas percebidas e pensadas [...] circunscrevendo [assim] nossas opções



práticas e cognitivas (DE SOUSA, 1987, p. 172).

Por exemplo: ter medo de um urso foca nossa atenção exclusivamente nas características da situação relevantes para uma fuga dela, sem que gastemos nosso tempo decidindo quais são os fatores irrelevantes a ignorar.

Damasio (1994, 2003) deu fundamentos neurobiológicos a essa ideia, sugerindo que as emoções simplificam o processo de decisão na medida em que demarcam rapidamente opções deliberativas no córtex pré-frontal como positivas ou negativas, à luz de suas consequências emocionais esperadas. Segundo Damasio, pacientes com dano pré-frontal ventromedial tornam-se irracionais como Hamlet quando diante de decisões banais como escolher uma data para a próxima consulta médica, irracionalmente propensos ao risco quando diante de decisões envolvendo apostas, e irracionalmente impacientes quando diante de decisões exigindo um adiamento de gratificação. O debate sobre se as evidências empíricas dão suporte à "hipótese do marcador somático" de Damasio continua em aberto (*vide* DUNN, DALGLEISH; LAWRENCE, 2006; REIMANN; BECHARA, 2010; BEER, 2017).

Uma outra influente ideia sobre a racionalidade das emoções é a de que elas ajudam a resolver o *problema do comprometimento* (SCHELLING, 1960; HIRSHLEIFER, 1987; FRANK, 1988), que é o desafio de convencer potenciais parceiros de cooperação de que o outro cumprirá com suas promessas e ameaças mesmo quando fortes considerações de interesse próprio sugerem o contrário. Por exemplo: Frank (1988) descreve a expressão de simpatia como um mecanismo de convencer possíveis cooperadores de que agiremos honestamente em interações futuras, mesmo quando diante de tentações, e a expressão da raiva como um mecanismo de convencer potenciais cooperadores de que agiremos agressivamente se provocados, mesmo quando a agressão nos for custosa. Esses sinais emocionais são tomados como críveis porque são difíceis de serem fingidos, e acabam beneficiando ambos os parceiros já que ajudam a estabelecer a disposição de ambas as partes a cooperar honestamente em empreitadas de benefício mútuo (*vide* ROSS; DUMOUCHEL, 2004; O'CONNOR, 2016).

## 11. Considerações finais

Desse panorama das teorias das emoções em diferentes disciplinas alguém poderia sentir-se tentado a dizer que a área está profundamente dividida quanto a

praticamente tudo; mas isso seria apressado. Apesar da grande diversidade de pontos de vista sobre a natureza e a função das emoções, aqui levantados, um amplo consenso emergiu sobre uma série de assuntos. Eis aqui uma possível lista daquilo sobre o que os teóricos das emoções concordam, com uma breve menção onde os desacordos começam:

- 1 Episódios emocionais envolvem, ao menos em casos prototípicos, um conjunto de componentes avaliativos, fisiológicos, fenomenológicos, expressivos, comportamentais e mentais que são indicativos das emoções e, em algum grau, correlacionam-se.

O grau em que estas correlações se dão continua a ser um tópico central do debate teórico: modelos de variável latente consideram que as emoções causam mudanças nos componentes e preveem fortes correlações, enquanto modelos de variável emergente consideram que as emoções emergem de mudanças nos componentes causadas por algo outro que as emoções, esperando encontrar correlações fracas.

- 2 Episódios ocorrentes do mesmo tipo de emoção popular (ex.: raiva, medo, vergonha) exibem uma grande variação de suas características expressivas, comportamentais, fisiológicas e fenomenológicas, bem como de sua intensidade, duração, valência, excitação, tipo e leque de objetos intencionais.

Pesquisadores discordam sobre se por trás de toda essa variação há padrões corporais mensuráveis de algum tipo que ainda fossem distintivos das diferentes emoções.

- 3 Emoções possuem intencionalidade ou a habilidade de representar.

Pesquisadores discordam sobre se as emoções representam descritivamente ou imperativamente ou ambos, sobre exatamente quais conteúdos elas representam, e sobre em que bases se dá a relação de representação emoção-mundo. Uma pequena

parcela de pesquisadores, oriundos sobretudo do movimento enativista, afirma que as emoções não possuem qualidades representacionais.

- 4 A base física das emoções é o cérebro, mas não há circuitos neurais que correspondam univocamente a nenhum tipo de emoção popular, e os cérebros são incorporados e situados em ambientes essenciais ao seu funcionamento adequado.

Pesquisadores discordam sobre como exatamente o cérebro implementa ocorrências de diferentes tipos de emoção, e sobre se os fenômenos emocionais são melhor compreendidos em termos de mecanismos neurais específicos ou não-específicos a emoções.

- 5 Emoções tipicamente incluem experiências conscientes, mas tais experiências não são estritamente necessárias para instanciar uma emoção, em parte porque alguns termos de emoções referem-se a disposições e em parte porque a maior parte dos teóricos considera as sensações [*feelings*] como conceitualmente distintas de emoções não-disposicionais.

Alguns pesquisadores influentes, como LeDoux (2017) e Barrett (2017), continuam a identificar as emoções como experiências conscientes.

- 6 Considerações evolucionárias e socioculturais devem contribuir para nosso entendimento das várias funções das emoções. Essas são funções intrapessoais, por exemplo: ajudar os organismos a coordenar recursos orgânicos na lida com demandas urgentes, e interpessoais, como comunicar informações úteis às transações sociais.

Pesquisadores continuam a debater se há evidências empíricas suficientes para as emoções básicas e outros mecanismos emocionais com propósitos especiais. Alguns veem o papel da evolução como limitado a moldar as adaptações com propósitos genéricos, como o afeto-central e a habilidade de

categorizar, que juntas levam à emergência das emoções.

7 Emoções não mais são consideradas como estruturalmente opostas à razão.

Pesquisadores continuam a debater as circunstâncias em que as emoções exibem diferentes tipos de irracionalidade cognitiva e estratégica.

8 Emoções podem ser adequadas e inadequadas no que diz respeito a seus objetos intencionais.

Pesquisadores debatem os fundamentos das, e as distinções entre, as diferentes formas de adequabilidade (ex.: cabimento, adequação moral).

9 Emoções tipicamente incluem apreciações da significância da situação de estímulo, variando entre formas primitivas e sofisticadas de processamento de informação.

Pesquisadores debatem o que são as estruturas de apreciação, e se elas causam ou constituem as emoções – ou ambas.

10 Emoções tipicamente correlacionam-se com mudanças na motivação para fazer coisas.

Alguns pesquisadores acreditam que as emoções causam ou consistem em tais mudanças de motivação, enquanto outros creem que as mudanças de motivação possuem outras causas, ou são vagas demais para fundar uma teoria do que as emoções são.

A exploração dessas ideias e a solução dos desacordos em torno delas é uma pujante empreitada interdisciplinar na teoria contemporânea da emoção. Filósofos e cientistas dos afetos ainda continuarão a trabalhar nela por anos, colocando suas distintivas habilidades teóricas a serviço de projetos de interesse comum.

## Referência bibliográfica

- ALLPORT, F. H. **Social Psychology**. Boston: Social Psychology, 1924.
- ANDREW, O.; TURNER, T. J. What's Basic About Basic Emotions?. **Psychological Review**, v. 97, n. 3, p. 315-331, 1990.
- ARMON-JONES, C. The Thesis of Constructionism. *In*: HARRÉ, R. (ed.). **The Social Construction of Emotions**. Oxford: Blackwell, 1986. p. 32-56.
- ARNOLD, M. B. **Emotion and Personality**. New York: Columbia University Press, 1960.
- ARPALY, N.; SCHRODER, T. Praise, Blame and the Whole Self. **Philosophical Studies**, v. 93, n. 2, p. 161-188, 2000.
- AVERILL, J. R. Emotion & Anxiety: Sociocultural, Biological and Psychological Determinants. *In*: RORTY, A. O. (ed.). **Explaining Emotions**. Los Angeles: University of California Press, 1980a. p. 37-72.
- AVERILL, J. R. Inner Feelings, Works of the Flesh, the Beast within, Diseases of the Mind, Driving Force, and Putting on a Show: Six Metaphors of Emotion and their Theoretical Extensions. *In*: LEARY, D. E. (ed.). **Metaphors in the History of Psychology**. New York: Cambridge University Press, 1990. p. 104-132.
- BAR-ON, D. Origins of Meaning: Must We "Go Gricean?". **Mind & Language**, v. 28, n. 3, p. 342-375, 2013.
- BARRETT, L. F. Feeling is Perceiving: Core Affect and Conceptualization in the Experience of Emotion. *In*: BARRETT, L. F.; NIEDENTHAL, P. M.; WINKIELMAN, P. (ed.). **Emotion and Consciousness**. New York: Guilford Press, 2005. p. 255-284.
- BARRETT, L. F. Are Emotions Natural Kinds?. **Perspectives on Psychological Science**, v. 1, n. 1, p. 28-58, 2006.
- BARRETT, L. F. Psychological Construction: The Darwinian Approach to the Science of Emotion. **Emotion Review**, v. 5, n. 4, p. 379-389, 2013.
- BARRETT, L. F. Ten Common Misconceptions about the Psychological Construction of Emotion. *In*: BARRETT, L. F.; RUSSELL, J. A. (ed.). **The Psychological Construction of Emotion**. New York: Guilford Press, 2015. p. 45-79.
- BARRETT, L. F. **How Emotions are Made: The Secret Life of The Brain**. New York, NY: Houghton-Mifflin-Harcourt, 2017.
- BARRETT, L. F.; LEWIS, M.; HAVILAND-JONES, J. M. (ed.). **Handbook of Emotions**. New York, NY: Guilford Press, 2016.

- BARRETT, L. F.; RUSSELL, J. A. (ed.). **The Psychological Construction of Emotion**. New York: Guilford Press, 2015.
- BARRETT, L. F.; SATPUTE, A. B. Large-Scale Brain Networks in Affective and Social Neuroscience: Towards an Integrative Architecture of the Human Brain. **Current Opinion in Neurobiology**, v. 23, n. 3, p. 361-372, 2013.
- BEDFORD, E. Emotions. **Proceedings of the Aristotelian Society**, v. 57, p. 281-304, 1957.
- BEER, J. S. Current Emotion Research in Social Neuroscience: How Does Emotion Influence Social Cognition?. **Emotion Review**, v. 9, n. 2, p. 172-180, 2017.
- BENBAJI, H. How is Recalcitrant Emotion Possible? **Australasian Journal of Philosophy**, v. 91, n. 3, p. 577-599, 2012.
- BEN-ZE'EV, A. **The Subtlety of Emotions**. Cambridge, MA: MIT Press, 2000.
- BRADY, M. S. The Irrationality of Recalcitrant Emotions. **Philosophical Studies**, v. 145, n. 3, p. 413-430, 2009.
- BRADY, M. S. **Emotional Insight: The Epistemic Role of Emotional Experience**. New York: Oxford University Press, 2013.
- BRENTANO, F. **Psychologie vom empirischen Standpunkte**. Leipzig: Duncker & Humblot. Traduzido como **Psychology from an Empirical Standpoint**. London: Routledge & K. Paul, 1973. Reimpresso com uma introdução de SIMONS, P. London: Routledge, 1995.
- BROAD, C. D. Emotion and Sentiment. **Journal of Aesthetics and Art Criticism**, v. 13, n. 2, p. 203-214, 1954. Reimpresso em CHENEY, D. R. (ed.). **Broad's Critical Essays in Moral Theory**. London: Allen & Unwin, 1971. DOI:10.2307/425913.
- BRUN, U. D.; DOĞUOĞLU, U.; KUENZLE, D. (ed.). **Epistemology and Emotions**. Aldershot: Ashgate, 2008.
- CANNON, W. B. **Bodily Changes in Pain, Hunger, Fear and Rage**. New York: Appleton, 1929.
- CARNAP, R. **Logical Foundations of Probability**. Chicago: University of Chicago Press, 1950.
- CARR, E. W. et al. Embodiment of Emotion and Its Situated Nature. *In*: NEWEN, A.; BRUIN, L.; GALLAGHER, S. (ed.). **The Oxford Handbook of 4E Cognition**. Oxford: Oxford University Press, 2018, p. 529-551.
- CHARLAND, L. C. Feeling and Representing: Computational Theory and the Modularity of Affect. *Synthese*, v. 105, n. 3, p. 273-301, 1995.

- CHARLAND, L. C. The Natural Kind Status of Emotion. *British Journal for the Philosophy of Science*, v. 53, n. 4, p. 511-537, 2002.
- CHOLBI, M. Grief's Rationality, Backward and Forward. *Philosophy and Phenomenological Research*, v. 94, n. 2, p. 255-272, 2017.
- CLARK, A. *Mindware: An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press, 2001.
- CLARK, J. A. Relations of Homology between Higher Cognitive Emotions and Basic Emotions. *Biology & Philosophy*, v. 25, n. 1, p. 75-94, 2010.
- CLARK-POLNER, E.; JOHNSON, T. D.; BARRETT, L. F. Multivoxel Pattern Analysis Does Not Provide Evidence to Support the Existence of Basic Emotions. *Cerebral Cortex*, v. 27, n. 3, p. 1944-1948, 2016.
- COLOMBETTI, G. *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: MIT Press, 2014.
- COLOMBETTI, G. Enactive Affectivity, Extended. *Topoi*, v. 36, n. 3, p. 445-455, 2017.
- COLOMBETTI, G.; ROBERTS, T. Extending the Extended Mind: The Case for Extended Affectivity. *Philosophical Studies*, v. 172, n. 5, p. 1243-1263, 2015.
- COLOMBETTI, G.; THOMPSON, E. The Feeling Body: Towards an Enactive Approach to Emotion. In: OVERTON, W. F.; MÜLLER, U.; NEWMAN, J. L. (ed.). *Developmental Perspectives on Embodiment and Consciousness*. New York: Erlbaum, 2008. p. 45-68.
- COVA, F.; FRIEND, S.; FORTHCOMING. How Does Fiction Elicit Emotions. In: SCARANTINO, A. (ed.). *Routledge Handbook of Emotion Theory*. Routledge, no prelo.
- DAMASIO, A. R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam's Sons, 1994.
- DAMASIO, A. R. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Orlando, FL: Harcourt, 2003.
- D'ARMS, J. The Moralistic Fallacy: On the "Appropriateness" of Emotion. *Philosophy and Phenomenological Research*, v. 61, p. 65-90, 2000.
- D'ARMS, J.; JACOBSON, D. The Significance of Recalcitrant Emotion (or, Anti-quasijudgmentalism). *Royal Institute of Philosophy Supplement*, v. 52, p. 127-45, 2003.
- DARWIN, C. **The Expression of the Emotions in Man and Animals**. Introduction, Notes and Commentaries by Paul Ekman. London: Harper Collins, 1872.
- DAVIDSON, D. How is Weakness of Will Possible?. In: FEINBERG, J. (ed.). **Moral Concepts**. Oxford: Oxford University Press, 1970. p. 21-43.
- DEIGH, J. Cognitivism in the Theory of Emotions. **Ethics**, v. 104, n. 4, p. 824-854, 1994.

- DELANCEY, C. **Passionate Engines: What Emotions Reveal about Mind and Artificial Intelligence**. Oxford: Oxford University Press, 2002.
- DEONNA, J. A. Emotions as Attitudes. **Dialectica**, v. 69, n. 3, p. 293-311, 2015.
- DEONNA, J. A.; TERONI, F. **The Emotions: A Philosophical Introduction**. London: Routledge, 2012.
- DEWEY, J. The Theory of Emotion. (1) Emotional Attitudes. **Psychological Review**, v. 1, n. 6, p. 553-569, 1894.
- DEWEY, J. The Theory of Emotion. (2) The Significance of Emotions. **Psychological Review**, v. 2, n. 1, p. 13-32, 1895.
- DIXON, T. **From Passions to Emotions: The Creation of a Secular Psychological Category**. Cambridge: Cambridge University Press, 2003.
- DIXON, T. "Emotion": the History of a Keyword in Crisis. **Emotion Review**, v. 4, n. 4, p. 338-344, 2012.
- DOKIC, J.; LEMAIRE, S. Are Emotions Perceptions of Value?. **Canadian Journal of Philosophy**, v. 43, n. 2, p. 227-247, 2013.
- DÖRING, S. A. Seeing What to Do: Affective Perception and Rational Motivation. **Dialectica**, v. 61, n. 3, p. 363-394, 2007.
- DÖRING, S. A. Conflict Without Contradiction. *In*: BRUN, U. D.; DOĞUOĞLU, U.; KUENZLE, D. (ed.). **Epistemology and Emotions**. Aldershot: Ashgate, 2008. p. 83-104.
- DÖRING, S. A.; LUTZ, A. Beyond Perceptualism: Introduction to the Special Issue. **Dialectica**, v. 69, n. 3, p. 259-270, 2015.
- DRETSKE, F. **Explaining Behavior**. Cambridge, MA: Bradford/MIT, 1988.
- DRETSKE, F. **Naturalizing the Mind**. Cambridge, MA: MIT Press, 1995.
- DUNN, B. D.; DALGLEISH, T.; LAWRENCE, A. D. The Somatic Marker Hypothesis: A Critical Evaluation. **Neuroscience and Biobehavioral Reviews**, v. 30, n. 2, p. 239-271, 2006.
- DURÁN, J.; REISENZEIN, R.; FERNÁNDEZ-DOLS, J.-M. Coherence between Emotions and Facial Expressions: A Research Synthesis. *In*: FERNÁNDEZ-DOLS, J.-M.; RUSSELL, J. A. (ed.). **The Science of Facial Expression**. New York: Oxford University Press, 2017. p. 107-129.
- EDER, A. B.; ROTHERMUND, K. Emotional Action: An Ideomotor Model. *In*: MOHIYEDDINI, C.; EYSENCK, M.; BAUER, S. (ed.). **Handbook of Psychology of Emotions: Recent Theoretical Perspectives and Novel Empirical Findings**. Hauppauge, NY: Nova Science Publishers, 2013.



- EKMAN, P. Biological and Cultural Contributions to Body and Facial Movement in the Expression of Emotions. RORTY, A. O. (ed.). **Explaining Emotions**. Los Angeles: University of California Press, 1980a. p. 73-102.
- EKMAN, P. Basic Emotions. *In*: DALGLEISH, T.; POWER, M. J. (ed.). **Handbook of Cognition and Emotion**. Chichester, UK: Wiley; Sons, 1999a. p. 45-60.
- EKMAN, P. Facial Expressions. *In*: DALGLEISH, T.; POWER, M. J. (ed.). **Handbook of Cognition and Emotion**. Chichester, UK: Wiley; Sons, 1999b. p. 301-320.
- EKMAN, P. **Emotions Revealed**. New York: Times Books, 2003.
- EKMAN, P.; CORDARO, D. What is Meant by Calling Emotions Basic. **Emotion Review**, v. 3, n. 4, p. 364-370, 2011.
- EKMAN, P.; FRIESEN, W. V. The Repertoire of Nonverbal Behavior. **Semiotica**, v. 1, n. 1, p. 86-88, 1969.
- EKMAN, P.; FRIESEN, W. V.; ELLSWORTH, P. **Emotion in the Human Face: Guidelines for Research and an Integration of Findings**. New York: Pergamon Press, 1972.
- ELGIN, C. Z. Emotion and Understanding. *In*: BRUN, U. D.; DOĞUOĞLU, U.; KUENZLE, D. (ed.). **Epistemology and Emotions**. Aldershot: Ashgate, 2008. p. 33-49.
- ELLSWORTH, P. C.; SCHERER, K. R. Appraisal Processes in Emotion. *In*: DAVIDSON, R. J.; SCHERER, K. R.; GOLDSMITH, H. (ed.). **Handbook of Affective Sciences**. Oxford: Oxford University Press, 2003. p. 572-595.
- ELSTER, J. **Alchemies of the Mind: Rationality and the Emotions**. Cambridge: Cambridge University Press, 1999.
- ELSTER, J. Emotional Choice and Rational Choice. *In*: GOLDIE, P. (ed.). **The Oxford Handbook of Philosophy of Emotion**. Oxford: Oxford University Press, 2010. p. 263-281.
- EVANS, D. **Emotion: The Science of Sentiment**. Oxford: Oxford University Press, 2001.
- EVANS, D.; CRUSE, P. (ed.). **Emotion, Evolution, and Rationality**. Oxford: Oxford University Press, 2004.
- FAUCHER, L. Facts and Values in Emotional Plasticity. *In*: CHARLAND, L.; ZACHAR, P. (ed.). **Fact and Value in Emotion**. Amsterdam: John Benjamins, 2008. p. 101-137.
- FAUCHER, L.; TAPPOLET, C. Introduction: Modularity and the Nature of Emotions. **Canadian Journal of Philosophy**, v. Supplementary Volume 32, 2008.
- FEHR, B.; RUSSELL, J. A. Concept of Emotion Viewed from a Prototype Perspective. **Journal of Experimental Psychology: General**, v. 113, n. 3, p. 464-486, 1984.

- FINGARETTE, H. **Self-Deception**. London: Routledge; Kegan Paul, 1969.
- FRANK, R. H. **Passions Within Reason: The Strategic Role of Emotions**. New York: Norton, 1988.
- FRIDLUND, A. J. **Human Facial Expression: An Evolutionary View**. San Diego, CA: Academic Press, 1994.
- FRIJDA, N. H. **The Emotions**. Cambridge: Cambridge University Press, 1986.
- FRIJDA, N. H. The Place of Appraisal in Emotion. **Cognition and Emotion**, v. 7, n. 3-4, p. 357-387, 1993.
- FRIJDA, N. H. Varieties of Affect: Emotions and Episodes, Moods and Sentiments. *In*: EKMAN, P.; DAVIDSON, R. J. (ed.). **The Nature of Emotion: Fundamental Questions**. New York: Oxford University Press, 1994. p. 59-67.
- FRIJDA, N. H. Impulsive Action and Motivation. **Biological Psychology**, v. 84, n. 3, p. 570-579, 2010.
- GALLAGHER, S. **Enactivist Interventions: Rethinking the Mind**. Oxford: Oxford University Press, 2017.
- GALLISTEL, C. R. **The Organization of Action: A New Synthesis**. Hillsdale, NJ: Erlbaum, 1980.
- GIBSON, J. J. **The Ecological Approach to Visual Perception**. Boston: Houghton Mifflin, 1979.
- GOLDIE, P. **The Emotions: A Philosophical Exploration**. Oxford: Oxford University Press, 2000.
- GOLDIE, P. Emotions, Feelings and Intentionality. **Phenomenology and the Cognitive Sciences**, v. 1, n. 3, p. 235-254, 2002.
- GOLDIE, P. Emotion, Feeling, and Knowledge of the World. *In*: SOLOMON, R. C. (ed.). **Thinking About Feelings: Philosophers on Emotions, (Series in Affective Science)**. Oxford; New York: Oxford University Press, 2003. p. 91-106.
- GOLDIE, P. Narrative, Emotion and Perspective. *In*: KIERAN, M.; LOPES, D. M. (ed.). **Imagination, Philosophy and the Arts**. London: Routledge, 2003b. p. 55-69.
- GOLDIE, P. Emotion, Reason and Virtue. *In*: EVANS, D.; CRUSE, P. (ed.). **Emotion, Evolution, and Rationality**. Oxford: Oxford University Press, 2004. p. 249-268.
- GOLDIE, P. (ed.). **The Oxford Handbook of Philosophy of Emotion**. Oxford: Oxford University Press, 2010.
- GOLDIE, P. **The Mess Inside: Narrative, Emotion, and the Mind**. Oxford: Oxford University Press, 2012.

- GOLDSTEIN, I. Are Emotions Feelings? A Further Look at Hedonic Theories of Emotions. **Consciousness and Emotion**, v. 3, n. 1, p. 21-33, 2002.
- GORDON, R. M. **The Structure of Emotions: Investigations in Cognitive Philosophy**. Cambridge: Cambridge University Press, 1987.
- GREEN, M. S. **Self-Expression**. Oxford: Clarendon Press, 2007.
- GREEN, O. H. **The Emotions: A Philosophical Theory**. Dordrecht, The Netherlands: Kluwer, 1992.
- GREENSPAN, P. Emotional Strategies and Rationality. **Ethics**, v. 110, n. 3, p. 469-487, 2000.
- GRIFFITHS, P. E. **What Emotions Really Are: The Problem of Psychological Categories**. Chicago: University of Chicago Press, 1997.
- GRIFFITHS, P. E. Towards a "Machiavellian" Theory of Emotional Appraisal. *In*: EVANS, D.; CRUSE, P. (ed.). **Emotion, Evolution, and Rationality**. Oxford: Oxford University Press, 2004. p. 89-105. DOI: 10.1093/acprof:oso/9780198528975.003.0005.
- GRIFFITHS, P.; SCARANTINO, A. Emotions in the Wild: the Situated Perspective on Emotion. *In*: ROBBINS, P.; AYDEDE, M. (ed.). **The Cambridge Handbook of Situated Cognition**. Cambridge: Cambridge University Press, 2009. p. 437-453. DOI: 10.1017/CBO9780511816826.023.
- GUNTHER, Y. H. Emotion and Force. *In*: GUNTHER, Y. H. (ed.). **Essays on Nonconceptual Content**. Cambridge, MA: MIT Press, 2003. p. 279-288.
- GUNTHER, Y. H. The Phenomenology and Intentionality of Emotion. **Philosophical Studies**, v. 117, n. 1-2, p. 43-55, 2004.
- GUSTAFSON, D. Grief. **Noûs**, v. 23, n. 4, p. 457-479, 1989.
- HARRÉ, R. (ed.). **The Social Construction of Emotions**. Oxford: Blackwell, 1986.
- HELM, B. W. **Emotional Reason: Deliberation, Motivation, and the Nature of Value**. Cambridge: Cambridge University Press, 2001.
- HELM, B. W. Emotions as Evaluative Feelings. **Emotion Review**, v. 1, n. 3, p. 248-255, 2009.
- HELM, B. W. Emotions and Recalcitrance: Reevaluating the Perceptual Model. **Dialectica**, v. 69, n. 3, p. 417-433, 2015.
- HESS, U.; BANSE, R.; KAPPAS, A. The Intensity of Facial Expressions is Determined by Underlying Affective State and Social Situation. **Journal of Personality and Social Psychology**, v. 69, n. 2, p. 280-288, 1995.
- HIRSHLEIFER, J. On the Emotions as Guarantors of Threats and Promises. *In*: DUPRÉ, J. (ed.). **The Latest on the Best: Essays on Evolution and**

- Optimality**. Cambridge, MA: MIT Press, 1987. p. 307-26.
- HORGAN, T.; TIENSON, J. The Intentionality of Phenomenology and the Phenomenology of Intentionality. *In*: CHALMERS, D. J. (ed.). **Philosophy of Mind: Classical and Contemporary Readings**. Oxford: Oxford University Press, 2002. p. 520-533.
- HUFENDIEK, R. **Embodied Emotions: A Naturalist Approach to a Normative Phenomenon**. London: Routledge, 2016.
- HUFENDIEK, R. Affordances and the Normativity of Emotions. **Synthese**, v. 194, n. 11, p. 4455-4476, 2017.
- HUFENDIEK, R. Explaining Embodied Emotions - With and Without Representations. **Philosophical Explorations**, v. 21, n. 2, p. 319-331, 2018.
- HURLEY, S. L. **Consciousness in Action**. Cambridge, MA: Harvard University Press, 1998.
- HURSTHOUSE, R. Arational Actions. **Journal of Philosophy**, v. 88, n. 2, p. 57-68, 1991.
- HUSSERL, E. Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Zweites Buch: Phänomenologische Untersuchungen zur Konstitution. The Hague: Martinus Nijhoff, 1989. Traduzido como: ROJCEWICZ, R.; SCHUWER, A. **Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: Second Book: Studies in the Phenomenology of Constitution**. Dordrecht: Kluwer, 1989.
- HUTTO, D. D. Truly Enactive Emotion. **Emotion Review**, v. 4, n. 2, p. 176-181, 2012.
- HUTTO, D. D.; MYIN, E. **Radicalizing Enactivism: Basic Minds without Content**. Cambridge, MA: MIT Press, 2013.
- IZARD, C. E. **The Face of Emotion**. New York: Appleton-Century-Crofts, 1971.
- IZARD, C. E. **Human Emotions**. New York: Plenum, 1977.
- IZARD, C. E. Basic Emotions, Relations Among Emotions, and Emotion-Cognition Relations. **Psychological Review**, v. 99, n. 3, p. 561-565, 1992.
- IZARD, C. E. Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. **Perspectives on Psychological Science**, v. 2, n. 3, p. 260-275, 2007.
- IZARD, C. E. Forms and Functions of Emotions: Matters of Emotion-Cognition Interactions. **Emotion Review**, v. 3, n. 4, p. 371-378, 2011.
- JAMES, W. What is an Emotion? **Mind**, v. 9, n. 2, p. 188-205, 1884.
- JAMES, W. **The Principles of Psychology**. New York: Holt, 1890.
- JOHNSTON, M. The Authority of Affect. **Philosophy and Phenomenological Research**, v. 63, n. 1, p. 181-214, 2001.

- KAGAN, J. **What is Emotion? History, Measures, and Meanings**. New Haven, CT: Yale University Press, 2007.
- KAGAN, J. Once More into the Breach. **Emotion Review**, v. 2, n. 2, p. 91-99, 2010.
- KELTNER, D.; HAIDT, J. Social Functions of Emotions. *In*: MAYNE, T. J.; BONANNO, G. A. (ed.). **Emotions: Current Issues and Future Directions**. New York: Guilford Press, 2001. p. 192-214.
- KELTNER, D. *et al.* **Expression of Emotion**. *In*: BARRETT, L. F.; LEWIS, M.; HAVILAND-JONES, J. M. (ed.). *In*: **Handbook of Emotions**. New York, NY: Guilford Press, 2016. p. 467-482.
- KENNY, A. **Action, Emotion and Will**. London, New York: Routledge; Kegan Paul; Humanities Press, 1963.
- KING, P. Emotions in Medieval Thought. *In*: GOLDIE, P. (ed.). **The Oxford Handbook of Philosophy of Emotion**. Oxford: Oxford University Press, 2010. p. 167-187.
- KLEEF, V.; A, G. **The Interpersonal Dynamics of Emotion: Toward an Integrative Theory of Emotions as Social Information**. Cambridge: Cambridge University Press, 2016.
- KLINEBERG, O. **Social Psychology**. New York: Holt, 1940.
- KOVACH, A.; LANCEY, C. D. On Emotions and the Explanation of Behavior. **Noûs**, v. 39, n. 1, p. 106-122, 2005.
- KRAGEL, P. A.; LABAR, K. S. Decoding the Nature of Emotion in the Brain. **Trends in Cognitive Sciences**, v. 20, n. 6, p. 444-455, 2016.
- KRAUT, R. Love De Re. **Midwest Studies in Philosophy**, v. 10, p. 413-430, 1986.
- KREIBIG, S. D. Autonomic Nervous System Activity in Emotion: A Review. **Biological Psychology**, v. 84, n. 3, p. 394-421, 2010.
- KRIEGEL, U. Towards a New Feeling Theory of Emotion. **European Journal of Philosophy**, v. 22, n. 3, p. 420-442, 2012.
- KRUEGER, J. Varieties of Extended Emotions. **Phenomenology and the Cognitive Sciences**, v. 13, n. 4, p. 533-555, 2014.
- LANDIS, C. Studies of Emotional Reactions: II. General Behavior and Facial Expression. **Journal of Comparative Psychology**, v. 4, n. 5, p. 447-509, 1924.
- LANGE, C. G. **Om sindsbevægelse: et psyko-fysiologisk Studie**. Traduzido como: HAUPT, A. **The Emotions**. Baltimore: Williams & Wilkins, 1885/1922.
- LAZARUS, R. S. Progress on a Cognitive-Motivational-Relational Theory of Emotion. **American Psychologist**, v. 46, n. 8, p. 819-834, 1991.
- LAZARUS, R. S. **Emotion and Adaptation**. New York: Oxford University Press, 1991b.

- LEDOUX, J. E. **The Emotional Brain: The Mysterious Underpinnings of Emotional Life**. New York: Simon; Schuster, 1996.
- LEDOUX, J. E. **Anxious: Using the Brain to Understand and Treat Fear and Anxiety**. New York: Viking, 2015.
- LEDOUX, J. E. Semantics, Surplus Meaning, and the Science of Fear. **Trends in Cognitive Science**, v. 21, n. 5, p. 303-306, 2017.
- LEVENSON, R. W. The Intrapersonal Function of Emotion. **Cognition and Emotion**, v. 13, n. 6, p. 481-504, 1999.
- LEVENSON, R. W. Basic Emotion Questions. **Emotion Review**, v. 3, n. 4, p. 379-438, 2011.
- LEWIS, M. D. Bridging Emotion Theory and Neurobiology through Dynamic Systems Modeling. **Behavioral and Brain Sciences**, v. 28, n. 2, p. 169-193, 2005.
- LEWIS, M.; HAVILAND-JONES, J. M.; BARRETT, L. F. (ed.). **Handbook of Emotions**. New York: Guilford Press, 2008.
- LINDQUIST, K. A. et al. The Brain Basis of Emotion: A Meta-Analytic Review. **Behavioral and Brain Sciences**, v. 35, n. 3, p. 121-202, 2012.
- LORMAND, E. Toward a Theory of Moods. **Philosophical Studies**, v. 47, n. 3, p. 385-407, 1985.
- LUTZ, C. **Emotion Words and Emotional Development on Ifaluk Atoll**. Tese. Harvard: Harvard University, 1980.
- LYONS, W. **Emotion**. Cambridge: Cambridge University Press, 1980.
- MANDLER, G. William James and the Construction of Emotion. **Psychological Science**, v. 1, n. 3, p. 179-180, 1990.
- MARKS, J. A Theory of Emotion. **Philosophical Studies**, v. 42, n. 2, p. 227-242, 1982.
- MATSUMOTO, D. et al. What's in a Face? Facial Expressions as Signals of Discrete Emotions. *In*: LEWIS, M.; HAVILAND-JONES, J. M.; BARRETT, L. F. (ed.). **Handbook of Emotions**. New York: Guilford Press, 2008. p. 211-234.
- MAUSS, I. B. et al. The Tie that Binds? Coherence Among Emotion Experience, Behavior, and Physiology. **Emotion**, v. 5, n. 2, p. 175-190, 2005.
- MCDUGALL, W. **An Introduction to Social Psychology**. Kitchener: Batoche Books, 1908.
- MCINTYRE, A. Is Akratic Action Always Irrational? *In*: FLANAGAN, O.; RORTY, A. O. (ed.). **Identity, Character, and Morality**. Cambridge, MA: MIT Press, 1990. p. 379-400
- MEINONG, A. Psychologisch-ethische Untersuchungen zur Werttheorie [Psychological-ethical investigations concerning the theory of value]. Graz: Leuschner,

1894. p. 3-244. Reimpresso em: HALLER, R.; KINDINGER, R. (ed.). **Alexius Meinong Gesamtausgabe Band III**. Graz: Akademische Druck—und Verlagsanstalt, 1968.
- MELE, A. R. **Irrationality: an Essay on Akrasia, Self-Deception, and Self-Control**. New York: Oxford University Press, 1987.
- MERLEAU-PONTY, M. **Phénoménologie de la perception**. Paris: Éditions Gallimard, 1945/1962. Traduzido como: *Phenomenology of Perception*. Tradução: Colin Smith. London: Routledge & K. Paul, 1945.
- MESQUITA, B.; BOIGER, M. Emotions in Context: A Sociodynamic Model of Emotions. **Emotion Review**, v. 6, n. 4, p. 298-302, 2014.
- MESQUITA, B.; FRIJDA, N. H. Cultural Variations in Emotions: A Review. **Psychological Bulletin**, v. 112, n. 2, p. 179-204, 1992.
- MESQUITA, B.; PARKINSON, B.; FORTHCOMING. Social Constructionist Theories of Emotions. In: SCARANTINO, A. (ed.). **Routledge Handbook of Emotion Theory**. Routledge, no prelo.
- MICELI, M.; CASTELFRANCHI, C. **Expectancy and Emotion**. Oxford: Oxford University Press, 2015.
- MOORS, A. On the Causal Role of Appraisal in Emotion. **Emotion Review**, v. 5, n. 2, p. 132-140, 2013.
- MOORS, A. Integration of Two Skeptical Emotion Theories: Dimensional Appraisal Theory and Russell's Psychological Construction Theory. **Psychological Inquiry**, v. 28, n. 1, p. 1-19, 2017.
- MOORS, A.; SCHERER, K. R. The Role of Appraisal in Emotion. In: ROBINSON, M.; WATKINS, E.; HARMON-JONES, E. (ed.). **Handbook of Cognition and Emotion**. New York: Guilford, 2013. p. 135-155.
- MORTON, A. Epistemic emotions. In: GOLDIE, P. (ed.). **The Oxford Handbook of Philosophy of Emotion**. Oxford: Oxford University Press, 2010. p. 385-400.
- MULLIGAN, K. From Appropriate Emotions to Values. **Monist**, v. 81, n. 1, p. 161-188, 1998.
- NEU, J. **A Tear is an Intellectual Thing: The Meanings of Emotion**. Oxford, New York: Oxford University Press, 2000.
- NIEDENTHAL, P. M. Embodying Emotion. **Science**, v. 316, p. 1002-05, 2007.
- NUMMENMAA, L.; SAARIMÄKI, H.; FORTHCOMING. Emotions as Discrete Patterns of Systemic Activity. **Neuroscience Letters**, published online, 10 jul. 2017.
- NUSSBAUM, M. C. **Upheavals of Thought: The Intelligence of Emotions**. Cambridge: Cambridge University Press, 2001.

- NUSSBAUM, M. C. **Anger and Forgiveness: Resentment, Generosity**. Justice, Oxford: Oxford University Press, 2016.
- OATLEY, K. **Best Laid Schemes: The Psychology of Emotions**. Cambridge: Cambridge University Press, 1992.
- OATLEY, K.; JOHNSON-LAIRD, P. Towards a Cognitive Theory of Emotions. **Cognition and Emotion**, v. 1, n. 1, p. 29-50, 1987.
- O'CONNOR, C. The Evolution of Guilt: a Model-Based Approach. **Philosophy of Science**, v. 83, n. 5, p. 897-908, 2016.
- O'REGAN, J.; NOË, A. A Sensorimotor Account of Vision and Visual Consciousness. **Behavioral and Brain Sciences**, v. 24, n. 5, p. 883-917, 2001.
- PANKSEPP, J. **Affective Neuroscience: the Foundations of Human and Animal Emotions**. New York, Oxford: Oxford University Press, 1998.
- PANKSEPP, J. Emotion as a Natural Kind within the Brain. *In*: LEWIS, M.; HAVILAND-JONES, J. M. (ed.). **Handbook of Emotions**. second ed. New York: Guilford University Press, 2000. p. 137-155.
- PAOLO, D.; EZEQUIEL; THOMPSON, E. The Enactive Approach. *In*: SHAPIRO, L. (ed.). **The Routledge Handbook of Embodied Cognition**. New York: Routledge, 2014. p. 68-78.
- PARKINSON, B. **Ideas and Realities of Emotion**. London; New York: Routledge, 1995.
- PARKINSON, B. Untangling the Appraisal-Emotion Connection. **Personality and Social Psychology Review**, v. 1, n. 1, p. 62-79, 1997.
- PARKINSON, B. Emotions in Direct and Remote Social Interaction: Getting through the Spaces between Us. **Computers in Human Behavior**, v. 24, n. 4, p. 1510-1529, 2008.
- PARKINSON, B. What Holds Emotions Together? Meaning and Response Coordination. **Cognitive Systems Research**, v. 10, n. 1, p. 31-47, 2009.
- PARKINSON, B.; FISCHER, A. H.; MANSTEAD, A. S. **Emotion in Social Relations: Cultural, Group, and Interpersonal Processes**. New York: Psychology Press, 2005.
- PARKINSON, B.; MANSTEAD, A. S. Appraisal as a Cause of Emotion. **Review of Personality and Social Psychology**, v. 13, p. 122-149, 1992.
- PICARD, R. W. **Affective Computing**. Cambridge: MIT Press, 1997.
- PITCHER, G. Emotion. **Mind**, v. 74, n. 295, p. 326-346, 1965.
- PLUTCHIK, R. **Emotion: A Psychoevolutionary Synthesis**. New York: Harper; Row, 1980.



- PRICE, C. Affect without Object: Moods and Objectless Emotions. **European Journal of Analytic Philosophy**, v. 2, n. 1, p. 49-68, 2006.
- PRINZ, J. **Gut Reactions: a Perceptual Theory of Emotion**. Oxford: Oxford University Press, 2004.
- PUGMIRE, D. **Rediscovering Emotion**. Edinburgh: Edinburgh University Press, 1998.
- PUGMIRE, D. **Sentiments: Integrity in the Emotions**. Edinburgh: Edinburgh University Press, 2005.
- RAMSEY, W. M. **Representation Reconsidered**. Cambridge: Cambridge University Press, 2007.
- RATCLIFFE, M. The Feeling of Being. **Journal of Consciousness Studies**, v. 12, n. 8-10, p. 43-60, 2005.
- RATCLIFFE, M. **Feelings of Being: Phenomenology, Psychiatry, and the Sense of Reality**. Oxford: Oxford University Press, 2008.
- RATCLIFFE, M. Grief and the Unity of Emotion. **Midwest Studies in Philosophy**, v. 41, p. 154-174, 2017.
- REIMANN, M.; BECHARA, A. The Somatic Marker Framework as a Neurological Theory of Decision-Making: Review, Conceptual Comparisons, and Future Neuroeconomics Research. **Journal of Economic Psychology**, v. 31, n. 5, p. 767-776, 2010.
- REISENZEIN, R. Emotional Action Generation. *In*: BATTMANN, W.; DUTKE, S. (ed.). **Processes of the Molar Regulation of Behavior**. Lengerich, Germany: Pabst Science, 1996. p. 151-165.
- REISENZEIN, R. Emotions as Metarepresentational States of Mind: Naturalizing the Belief-Desire Theory of Emotion. **Cognitive Systems Research**, v. 10, n. 1, p. 6-20, 2009a
- REISENZEIN, R. Emotional Experience in the Computational Belief-Desire Theory of Emotion. **Emotion Review**, v. 1, n. 3, p. 214-222, 2009b.
- ROBERTS, R. C. **Emotions: An Essay in Aid of Moral Psychology**. New York: Cambridge University Press, 2003.
- ROBINSON, J. **Deeper than Reason: Emotion and Its Role in Literature, Music and Art**. New York, Oxford: Oxford University Press, 2005.
- RORTY, A. O. (ed.). **Explaining Emotions**. Los Angeles: University of California Press, 1980a.
- RORTY, A. O. Explaining Emotions. *In*: RORTY, A. O. (ed.). **Explaining Emotions**. Los Angeles: University of California Press, 1980a. p. 103-126.

- RORTY, A. O. The Historicity of Psychological Attitudes: Love is Not Love Which Alters Not When it Alteration Finds. **Mind in Action: Essays in the Philosophy of Mind**, v. 10, n. 1, p. 3997-412, 1987.
- RORTY, A. O. Enough Already with Theories of Emotion. *In*: SOLOMON, R. C. (ed.). **Thinking About Feelings: Philosophers on Emotions, (Series in Affective Science)**. Oxford; New York: Oxford University Press, 2003. Cap. 17.
- ROSEMAN, I. J. Why These Appraisals? Anchoring Appraisal Models to Research on Emotional Behavior and Related Response Systems. *In*: FRIJDA, N. H. (ed.). **Proceedings of the Ninth International Conference of the International Society for Research on Emotions**. Toronto: International Society for Research on Emotions, 1996, p. 106-110.
- ROSEMAN, I. J.; SMITH, C. A. Appraisal Theory: Overview, Assumptions, Varieties, Controversies. *In*: SCHERER, K. R.; SCHORR, A.; JOHNSTONE, T. **Appraisal Processes in Emotion: Theory, Methods, Research**. Oxford: Oxford University Press, 2001. p. 3-34.
- ROSS, D.; DUMOUCHEL, P. Emotions as Strategic Signals. **Rationality and Society**, v. 16, n. 3, p. 251-286, 2004.
- ROSSI, M.; TAPPOLET, C.; FORTHCOMING. What Kind of Evaluative States Are Emotions? The Attitudinal Theory vs. the Perceptual Theory of Emotions. **Canadian Journal of Philosophy**, p. 1-20, jun. 2018.
- RUSSELL, J. A. A Circumplex Model of Affect. **Journal of Personality and Social Psychology**, v. 39, n. 6, p. 1161-1178, 1980.
- RUSSELL, J. A. Comment on Articles by Frijda and by Conway and Bekerian. **Cognition and Emotion**, v. 1, n. 2, p. 193-197, 1987. DOI: 10.1080/02699938708408045.
- RUSSELL, J. A. Culture and the Categorization of Emotions. **Psychological Bulletin**, v. 110, n. 3, p. 426-450, 1991. DOI: 10.1037/0033-2909.110.3.426.
- RUSSELL, J. A. Is There Universal Recognition of Emotion from Facial Expressions? A Review of the Cross-Cultural Studies. **Psychological Bulletin**, v. 115, n. 1, p. 102-141, 1994. DOI: 10.1037/0033-2909.115.1.102.
- RUSSELL, J. A. Core Affect and the Psychological Construction of Emotion. **Psychological Review**, v. 110, n. 1, p. 145-172, 2003. DOI: 10.1037/0033-295X.110.1.145.
- RUSSELL, J. A. From a Psychological Constructionist Perspective. *In*: ZACHAR, P.; ELLIS, R. D. (ed.). **Categorical Versus Dimensional Models of Affect: A Seminar on the Theories of Panksepp and Russell, (Consciousness**

- and Emotion book series**). Amsterdam, The Netherlands: John Benjamins, 2012. p. 79-118. DOI: 10.1075/ceb.7.03rus.
- SALMELA, M. Can Emotion be Modelled on Perception? *Dialectica*, v. 65, n. 1, p. 1-29, 2011.
- SARTRE, J.-P. **Esquisse d'une théorie des émotions**. Traduzido como: **The Emotions: Outline of a Theory**. Tradução: Bernard Frechtman. New York: Philosophical Library, 1939.
- SCARANTINO, A. Insights and Blindspots of the Cognitivist Theory of Emotions. *British Journal for the Philosophy of Science*, v. 61, n. 4, p. 729-768, 2010.
- SCARANTINO, A. How To Define Emotions Scientifically. *Emotion Review*, v. 4, n. 4, p. 358-368, 2012.
- SCARANTINO, A. The Motivational Theory of Emotions. In: DANIEL JACOBSON, J. D. (ed.). **Moral Psychology and Human Agency**. Oxford: Oxford University Press, 2014. p. 156-185.
- SCARANTINO, A. Basic Emotions, Psychological Construction and the Problem of Variability. In: BARRETT, L. F.; RUSSELL, J. A. (ed.). **The Psychological Construction of Emotion**. New York: Guilford Press, 2015. p. 334-376.
- SCARANTINO, A. The Philosophy of Emotions and Its Impact on Affective Science. In: BARRETT, L. F.; LEWIS, M.; HAVILAND-JONES, J. M. (ed.). **Handbook of Emotions**. New York, NY: Guilford Press, 2016. p. 3-48.
- SCARANTINO, A. How to Do Things with Emotional Expressions: The Theory of Affective Pragmatics. *Psychological Inquiry*, v. 28, n. 2-3, p. 165-185, 2017.
- SCARANTINO, A.; GRIFFITHS, P. Don't Give up on Basic Emotions. *Emotion Review*, v. 3, n. 4, p. 444-454, 2011.
- SCARANTINO, A.; NIELSEN, M. Voodoo Dolls and Angry Lions: How Emotions Explain Arational Actions. *Philosophical Studies*, v. 172, n. 11, p. 2975-2998, 2015.
- SCARANTINO, A. (ed.). **Routledge Handbook of Emotion Theory**. Routledge, no prelo.
- SCHACHTER, S.; SINGER, J. Cognitive, Social and Physiological Determinants of Emotional States. *Psychological Review*, v. 69, n. 5, p. 379-399, 1962.
- SCHELLING, T. C. **The Strategy of Conflict**. Cambridge, MA: Harvard University Press, 1960.
- SCHERER, K. R. On the Nature and Function of Emotion: A Component Process Approach. In: SCHERER, K. R.; EKMAN, P. (ed.). **Approaches to Emotion**. Hillsdale, NJ: Erlbaum, 1984.

- SCHERER, K. R. Appraisal Considered as a Process of Multilevel Sequential Checking. *In*: SCHERER, K. R.; SCHORR, A.; JOHNSTONE, T. **Appraisal Processes in Emotion: Theory, Methods, Research**. Oxford: Oxford University Press, 2001. p. 92-120.
- SCHERER, K. R. What are Emotions? And How Can They be Measured? **Social Science Information**, v. 44, n. 4, p. 695-729, 2005.
- SCHERER, K. R. The Dynamic Architecture of Emotion: Evidence for the Component Process Model. **Cognition and Emotion**, v. 23, n. 7, p. 1307-1351, 2009.
- SCHERER, K. R.; MOORS, A.; The Emotion Process: Event Appraisal and Component Differentiation. **Annual Review of Psychology**, v. 70, p. 719-745, 2019. Online first, 2018.
- SCHERER, K. R.; SCHORR, A.; JOHNSTONE, T. **Appraisal Processes in Emotion: Theory, Methods, Research**. Oxford: Oxford University Press, 2001.
- SCHWARZ, N. Feelings-As-Information Theory. *In*: LANGE, P. A.; KRUGLANSKI, A. W.; HIGGINS, E. (ed.). **Handbook of Theories of Social Psychology**. London: Sage Publications, 2012, p. 289-308.
- SEARLE, J. R. The Intentionality of Intention and Action. **Inquiry**, v. 22, n. 1-4, p. 253-280, 1979.
- SHAND, A. F. **The Foundations of Character: A Study of the Emotions and Sentiments**. London: Macmillan, 1920.
- SHARGEL, D. **Constitutively Embodied Emotions**. Tese. City University of New York, 2014. Disponível em: [https://academicworks.cuny.edu/gc\\_etds/472](https://academicworks.cuny.edu/gc_etds/472). Acesso em: 07 out. 2021.
- SHARGEL, D. Emotions without Objects. **Biology & Philosophy**, v. 30, n. 6, p. 831-844, 2015.
- SHARGEL, D.; PRINZ, J. An Enactivist Theory of Emotional Content. *In*: NAAR, H.; TERONI, F. (ed.). **The Ontology of Emotions**. Cambridge: Cambridge University Press, 2018. p. 110-129.
- SILVIA, P. J. **The Psychology of Interest**. New York: Oxford University Press, 2006.
- SLABY, J.; WÜSCHNER, P. Emotion and Agency. *In*: ROESER, S.; TODD, C. (ed.). **Emotion and Value**. Oxford: Oxford University Press, 2014. p. 212-228.
- SMITH, C. A.; ELLSWORTH, P. C. Patterns of Cognitive Appraisal in Emotion. **Journal of Personality and Social Psychology**, v. 48, n. 4, p. 813-838, 1985.
- SMITH, M. The Possibility of Philosophy of Action. *In*: BRANSEN, J.; CUYPERS, S. E. (ed.). **Human Action, Deliberation and Causation**. The Netherlands: Springer, 1998. p. 17-41.

- SOLOMON, R. C. **The Passions, Garden City**. New York: Doubleday Anchor, 1976.
- SOLOMON, R. C. (ed.). **Thinking About Feelings: Philosophers on Emotions**, (Series in Affective Science). Oxford; New York: Oxford University Press, 2003.
- SOLOMON, R. C. The Philosophy of Emotions. *In*: LEWIS, M.; HAVILAND-JONES, J. M.; BARRETT, L. F. (ed.). **Handbook of Emotions**. New York: Guilford Press, 2008. p. 3-16.
- SOLOMON, R. C. Emotions and Choice. *In*: RORTY, A. O. (ed.). **Explaining Emotions**. Los Angeles: University of California Press, 1980a. p. 251-81.
- SOUSA. Education and Time. **Metaphilosophy**, v. 21, p. 434-446, 1990.
- SOUSA. **Emotional Truth**. Oxford: Oxford University Press, 2011.
- SOUSA; RONALD. **The Rationality of Emotion**. Cambridge, MA: MIT Press, 1987.
- STEPHAN, A. Moods in Layers. **Philosophia**, v. 45, n. 4, p. 1481-1495, 2017.
- STEPHAN, A. On the Adequacy of Emotions and Existential Feelings. **Rivista Internazionale di Filosofia e Psicologia**, v. 8, n. 1, p. 1-13, [s.d.].
- STEPHAN, A.; WALTER, S.; WILUTZKY, W. Emotions Beyond Brain and Body. **Philosophical Psychology**, v. 27, n. 1, p. 98-111, 2014.
- STEPHENS, C. L.; CHRISTIE, I. C.; FRIEDMAN, B. H. Autonomic Specificity of Basic Emotions: Evidence from Pattern Classification and Cluster Analysis. **Biological Psychology**, v. 84, n. 3, p. 463-473, 2010.
- TAPPOLET, C. **Emotions et Valeurs**. Paris: Presses Universitaires de France, 2000.
- TAPPOLET, C. Emotions, Action, and Motivation: the Case of Fear. *In*: GOLDIE, P. (ed.). **The Oxford Handbook of Philosophy of Emotion**. Oxford: Oxford University Press, 2010. p. 325-345.
- TAPPOLET, C. **Emotions, Values, and Agency**. Oxford: Oxford University Press, 2016.
- TAYLOR, G. Justifying the Emotions. **Mind**, v. 84, n. 1, p. 390-402, 1975.
- TERONI, F. Emotions and Formal Objects. **Dialectica**, v. 61, n. 3, p. 395-415, 2007.
- THAGARD, P. **Hot Thought: Mechanisms and Applications of Emotional Cognition**. Cambridge, MA: MIT Press, 2006.
- THALBERG, I. Emotion and Thought. **American Philosophical Quarterly**, v. 1, n. 1, p. 45-55, 1964.
- THOMPSON, E. **Mind in Life: Biology, Phenomenology, and the Sciences of Mind**. Cambridge, MA: Harvard University Press, 2007.
- TOMKINS, S. S. **Affect Imagery Consciousness: The Complete Edition**. New York: Springer Publishing Company, 2008. v. vols. I, II, III, IV
- TOOBY, J.; COSMIDES, L. The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables. *In*: LEWIS, M.; HAVILAND-

- JONES, J. M.; BARRETT, L. F. (ed.). **Handbook of Emotions**. New York: Guilford Press, 2008. p. 114-137.
- Toward a Dialect Theory: Cultural Differences in the Expression and Recognition of Posed Facial Expressions. **Emotion**, v. 7, n. 1, p. 131-146, 2007.
- TYE, M. **Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind**. Cambridge, MA: MIT Press, 1995.
- UEXKÜLL, J. VON. Streifzüge durch die Umwelten von Tieren und Menschen, 1934. Traduzido como: **A Foray into the Worlds of Animals and Humans: With a Theory of Meaning**. Tradução: Joseph D. O'Neill. Minneapolis: University of Minnesota Press, 2010.
- VAN LEEUWEN, D. The Spandrels of Self-Deception. **Philosophical Psychology**, v. 20, n. 3, p. 329-348, 2007.
- VARELA, F. J. Neurophenomenology: A Methodological Remedy for the Hard Problem. **Journal of Consciousness Studies**, v. 3, n. 4, p. 330-350, 1996.
- VARELA, F. J.; THOMPSON, E.; ROSCH, E. **The Embodied Mind: Cognitive Science and Human Experience**. Cambridge, MA: MIT Press, 1991.
- WHITING, D. The Feeling Theory of Emotion and the Object-Directed Emotions. **European Journal of Philosophy**, v. 19, n. 2, p. 281-303, 2011.
- WIERZBICKA, A. **Emotions across Languages and Cultures: Diversity and Universals**. Cambridge: Cambridge University Press, 1999.
- WILSON-MENDENHALL, C. D.; LISA FELDMAN BARRETT, W. S.; BARSALOU, L. W. Grounding Emotion in Situated Conceptualization. **Neuropsychologia**, v. 49, n. 5, p. 1105-1127, 2011.
- WOLLHEIM, R. **On the Emotions**. New Haven: Yale University Press, 1999.
- ZACHAR, P. The Classification of Emotion and Scientific Realism. **Journal of Theoretical and Philosophical Psychology**, v. 26, n. 1-2, p. 120-138, 2006.
- ZACHAR, P.; ELLIS, R. D. (ed.). **Categorical Versus Dimensional Models of Affect: A Seminar on the Theories of Panksepp and Russell**, (Consciousness and Emotion book series). Amsterdam, The Netherlands: John Benjamins, 2012.
- ZAJONC, R. Feeling and Thinking: Closing the Debate over the Independence of Affect". *In*: FORGAS, J. P. (ed.). **Feeling and Thinking: The Role of Affect in Social Cognition**. Cambridge: Cambridge University Press, 2000. p. 31-58.
- ZINCK, A.; NEWEN, A. Classifying Emotion: A Developmental Account. **Synthese**, v. 161, n. 1, p. 1-25, 2008.

## (VII) Cognição Animal\*

Autor: Kristin Andrews

Tradução: Bruno Malavolta e Silva (PPGFIL-UFRGS)

Revisão: Filipe Lazzeri Vieira (UFG)

As questões filosóficas relacionadas à pesquisa sobre cognição animal podem ser categorizadas em três grupos: questões fundacionais sobre se os animais não humanos constituem um tema adequado para investigação psicológica; questões metodológicas sobre como estudar mentes animais; e questões mais específicas que surgem de dentro de programas de pesquisa específicos.

Embora o estudo da cognição animal seja em grande parte um empreendimento empírico, a prática da ciência nesta área é baseada em argumentos e suposições teóricas; por exemplo, sobre a natureza da mente, da comunicação e da racionalidade. Se animais não humanos não possuem crenças, e se todos os sistemas cognitivos

---

\*ANDREWS, K. Animal Cognition. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Summer Edition. Stanford, CA: The Metaphysics Research Lab, 2016. Disponível em: <https://plato.stanford.edu/archives/sum2016/entries/cognition-animal/>. Acesso em: 30 out. 2021.

The following is the translation of the entry on Animal Cognition by Akristin Andrews, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at <https://plato.stanford.edu/archives/sum2016/entries/cognition-animal/>. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. The current version is located at <https://plato.stanford.edu/entries/cognition-animal/>. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry. Finally, we would like to thank to John Templeton Foundation for financially supporting this project.

possuem crenças, então animais não constituem um tópico adequado para estudos cognitivos. Se animais não são agentes porque seu comportamento não é causado por atitudes proposicionais, e se todos os sistemas cognitivos são agentes, chegamos à mesma conclusão. Embora existam argumentos contrários a mentes animais, os cientistas cognitivos que estudam animais aceitam amplamente que animais são sistemas cognitivos que possuem mente. Conforme demonstrado pela Declaração de Cambridge sobre a Consciência feita em 2012, muitos cientistas também aceitam a consciência animal.

Muitos dos programas de pesquisa que investigam capacidades cognitivas particulares em diferentes espécies levantam questões filosóficas e possuem implicações para teorias filosóficas, na medida em que impõem restrições empíricas adicionais para filósofos naturalistas. Paradigmas de pesquisa tradicionais em cognição animal são similares aos de cognição humana, e incluem um exame da percepção, aprendizagem, categorização, memória, cognição espacial, numerosidade, comunicação, linguagem, cognição social, teoria da mente ou leitura de mente, raciocínio causal e metacognição.

Os cientistas que trabalham em qualquer uma dessas áreas podem usar métodos muito diferentes, porque não existe uma disciplina singular de cognição animal. Por exemplo, a cognição social pode ser estudada por um biólogo que documenta trocas de olhares entre pares formados por mãe e bebê em diferentes espécies de primatas (*vide* MATSUZAWA, 2006b), por um etólogo interessado no comportamento de brincadeiras sociais entre canídeos em liberdade (*vide* BEKOFF, 2001), por um psicólogo experimental testando teoria da mente em um chimpanzé adulto treinado com símbolos (*vide* PREMACK; WOODRUFF, 1978), por um antropólogo observando jogos sociais em comunidades de macacos-prego (*vide* PERRY *et al.*, 2003) ou por um neurocientista cognitivo que investiga a base neural para o rastreamento de olhares em primatas (*vide* EMERY, 2000). Por fim, as descobertas sobre as habilidades cognitivas de animais muitas vezes desempenham um papel em debates sobre o status moral dos animais, bem como nas investigações sobre se animais podem se envolver em algum tipo de prática moral.

## 1. O que é Cognição Animal?

A pesquisa em cognição animal examina os processos usados para gerar comportamento adaptativo ou flexível em espécies animais. Muito do trabalho sobre



cognição animal é mais apropriadamente descrito pela expressão cognição comparativa, porque os processos e capacidades subjacentes ao comportamento são comparados entre diferentes espécies (SHETTLEWORTH, 2010). No contexto da pesquisa sobre cognição animal, a pesquisa da cognição geralmente se concentra em questões sobre os mecanismos envolvidos em capacidades específicas, como aprendizagem, memória, percepção ou poder de decisão. Pesquisadores também investigam os conceitos, crenças e pensamentos de animais. Embora a teoria representacional da mente seja uma suposição comum entre pesquisadores de cognição animal, também há investigação sobre o papel que a percepção desempenha na aprendizagem animal, e interesse enquanto trabalho explicativo pode ser feito por conteúdo não conceitual, às vezes inspirado por trabalhos em cognição corporificada (BARRETT, 2011). E, embora os processos cognitivos sejam frequentemente contrastados com os processos associativos, essa distinção é frequentemente contestada (BUCKNER 2011; MITCHELL *et al.*, 2009). Como parte da ciência cognitiva, a pesquisa em cognição animal visa descobrir os diferentes mecanismos cognitivos atuantes entre as espécies, com o objetivo de compreender as variedades de cognição, as semelhanças entre os humanos e outras espécies, e a evolução e função dos processos cognitivos.

## 2. Questões Fundacionais

A discussão filosófica da cognição animal tem sido tradicionalmente focada na metafísica e epistemologia da mente em criaturas que não têm linguagem. Alguns filósofos se perguntaram se os animais têm mente, são racionais, têm conceitos e crenças; além de se engajarem na questão de como responder a essas perguntas diante das limitações inerentes à investigação.

O período inicial da história da filosofia ocidental reflete uma tendência de ver os animais como desprovidos de racionalidade. Aristóteles definiu “humano” como “o animal racional”, rejeitando assim a possibilidade de qualquer outra espécie ser racional (ARISTÓTELES, *Metafísica*). Tomás de Aquino acreditava que os animais são irracionais por não serem livres (TOMÁS DE AQUINO, *Summa Theologica*). Séculos depois, Descartes defendeu uma distinção entre humanos e animais com base na crença de que a linguagem é uma condição necessária para a mente racional; em sua visão, os animais são máquinas sem alma (DESCARTES, *Discurso do Método*). Locke concordou que os animais não podem pensar, porque

as palavras são necessárias para compreender os universais (LOCKE, *Ensaio sobre o Entendimento Humano*). Seguindo essa tradição, Kant concluiu que, por não poderem pensar sobre si mesmos, os animais não são agentes racionais e, por isso, possuem apenas valor instrumental (KANT, *Leituras sobre Ética*).

No entanto, também houve quem discordasse. Voltaire criticou a visão de Descartes, alegando que podemos ver as mentes dos animais, assim como vemos as mentes humanas, e é através da interação com os animais que passamos a julgar que eles têm emoções, memórias e crenças (VOLTAIRE, *Dicionário Filosófico*). Hume foi inteiramente desdenhoso aos céticos da mente animal quando escreveu: “Próximo ao ridículo de negar uma verdade evidente, está o de se esforçar muito para defendê-la; e nenhuma verdade me parece mais evidente do que a de que os animais são dotados de pensamento e razão tanto quanto o homem. Os argumentos são, neste caso, tão óbvios, que nunca escapam ao mais estúpido e ignorante” (HUME, *Tratado da Natureza Humana*, p. 176).

## 2.1 Mentes Animais

Apesar do julgamento de Hume sobre o valor de tais argumentos, muita tinta foi derramada em defesa da existência de mentes animais, o que é seguidamente visto como uma versão especial do problema de outras mentes. Três tipos de argumentos para outras espécies de mente são: o *argumento por analogia*, o *argumento da inferência pela melhor explicação*, e o *argumento da parcimônia evolucionária*.

O argumento por analogia para mentes animais pode ser formulado como:

- 1 Todos os animais que já sei que têm mente (ou seja, humanos) têm a propriedade *x*.
- 2 Os indivíduos da espécie *y* possuem a propriedade *x*.
- 3 Logo, os indivíduos da espécie *y* provavelmente têm mentes.

As versões do argumento diferem quanto ao que escolhem como a propriedade referida por *x*, e em como defendem a escolha dessa propriedade. Tal propriedade de referência pode referir a uma série de coisas, como a uma capacidade geral (por exemplo, de resolução de problemas), uma habilidade específica (por

exemplo, linguagem, teoria da mente, uso de ferramentas), um comportamento, ou até mesmo atividade cerebral. Confira Farah (2008) para um exemplo desta última abordagem.

O argumento por analogia para mentes animais é em certo sentido mais forte do que o argumento para outras mentes, na medida em que a classe de referência é maior. Em vez de começar com a introspecção de sua própria mente e então generalizar para todos os outros humanos, o argumento em favor das mentes animais considera que todos os humanos têm mentes e generaliza da espécie humana para outras espécies. No entanto, em outro sentido o argumento analógico para mentes animais é mais fraco, uma vez que a força do argumento é uma função do grau de semelhança entre a classe de referência e a classe alvo. Os humanos são provavelmente mais parecidos uns com os outros do que com membros de outras espécies. Embora alguns pesquisadores que trabalham com grandes símios tenham expressado preocupação com o argumento por analogia (POVINELLI *et al.*, 2000; POVINELLI, 2000), foi apontado que essas preocupações não têm por objeto precisamente os argumentos analógicos para a existência de mentes animais (ALLEN, 2002). Além disso, os usos recentes do argumento analógico para defender a consciência animal têm sido baseados em uma investigação cuidadosa dos poderes causais da propriedade de referência (VARNER, 2012).

O *argumento da inferência pela melhor explicação* repousa na afirmação de que a existência de mentes animais é uma explicação melhor para o comportamento e a fisiologia animal do que as oferecidas por outras hipóteses. Uma versão desse argumento pode ser formulada como:

- 1 Indivíduos da espécie *x* se envolvem em comportamentos *y*.
- 2 A melhor explicação científica para um indivíduo que se engaja em comportamentos *y* é a de que eles têm uma mente.
- 3 Portanto, é provável que indivíduos da espécie *x* tenham mentes.

O argumento da inferência pela melhor explicação justifica a atribuição de estados mentais aos animais com base no robusto poder preditivo e explicativo que é obtido de tais atribuições. Segundo o argumento, sem essas atribuições, seríamos incapazes de entender o comportamento animal. Este argumento se baseia no raciocínio científico ordinário; entre duas hipóteses, a que melhor explica o fenômeno

é tomada como preferível. Os proponentes deste tipo de argumento para mentes animais alegam que ter uma mente (seja lá o que isso for) explica melhor o comportamento observado.

Embora seja justo dizer que a maioria dos cientistas que trabalham com animais pensa que eles têm mentes, tais cientistas também tendem a usar a estratégia de inferência pela melhor explicação quando discordam sobre os mecanismos cognitivos em ação, em uma espécie particular em um momento particular.

O *argumento da parcimônia evolucionária* é baseado na ideia de que espécies intimamente relacionadas compartilham alguns traços físicos, e essa relação pode oferecer evidências em favor de uma explicação causal mentalista em certos casos. Este argumento pode ser formulado de forma diferente dependendo da noção de parcimônia (SOBER, 2015). Em geral, tais argumentos sugerem que o fato de compartilharmos alguma propriedade com um animal é suficiente para estabelecer o animal como (provavelmente) dotado de mente, se assumirmos (a) que compartilhamos um ancestral comum com o animal e (b) que devemos preferir a explicação mais parcimoniosa do surgimento dessa propriedade (DE WAAL, 1999; SOBER, 2005). No entanto, sem o conhecimento da mentalidade do ancestral comum, tais argumentos oferecem pouca evidência adicional (SOBER, 2014).

Outros filósofos contemporâneos sugerem, como Hume e Voltaire, que não precisamos de argumentos para mentes de outros animais, porque vemos as mentes diretamente quando estamos interagindo com elas. Enquanto abraçamos nossos cães e lutamos com nossos gatos, vemos as mentes de nossos companheiros animais tão claramente quanto vemos as mentes de nossos bebês quando brincamos com eles. Basear-se em um argumento para outras mentes sempre abre a possibilidade de que o argumento resulte ruim e a conclusão falsa. Mas não importa o argumento que encontrarmos, não seremos capazes de agir como se negássemos as mentes que vemos em cães, gatos e bebês humanos (JAMIESON, 1998, 2009; SEARLE, 1994).

## 2.2 Antropomorfismo e Psicologia do Senso Comum

Embora possamos concordar que os animais têm mentes, preocupações surgem quando começamos a descrever as características dessas mentes. Quando os pesquisadores atribuem conteúdo mental a outras espécies, eles se expõem à acusação de antropomorfismo. O termo “antropomorfismo” tem várias conotações diferentes, mas geralmente se refere ao ato de atribuir características humanas a

outros animais. Às vezes o termo é usado para se referir apenas a características psicológicas, e às vezes é usado para se referir a características que se afirmam serem exclusivamente humanas (nesse caso, o antropomorfismo é um erro por definição). Nos últimos anos, houve uma série de discussões teóricas sobre a acusação de antropomorfismo ela própria (DASTON; MITMAN, 2005; MITCHELL *et al.*, 1997; ANDREWS, 2009; ANDREWS; HUSS, 2014; ASQUITH, 1997; BUCKNER, 2013; CRIST, 1999; FISHER, 1990, 1991; KEELEY, 2004; KENNEDY, 1992; MEKETA, 2014; RIVAS; BURGHARDT, 2002; SHETTLEWORTH, 2010b; SOBER, 2012; WYNNE, 2004, 2007), incluindo preocupações mais específicas com o uso da psicologia de senso comum na pesquisa da cognição animal (PENN, 2011; ANDREWS, no prelo).

Uma maneira de ver um erro antropomórfico é como um erro de categoria, em vez de uma falsa atribuição. Um erro antropomórfico deve ser logicamente falso, porque os membros das espécies que se tem em vista não são o tipo de coisa a que o termo pode ser aplicado (KEELEY, 2004; FISHER, 1991). No entanto, se a acusação de antropomorfismo é a acusação de que o atribuidor está cometendo um erro de categoria, então a acusação está sendo feita em bases conceituais, em vez de empíricas. Assim, uma resposta à acusação de antropomorfismo é a pesquisa contínua, pois não se saberia se uma propriedade é antropomórfica até que a pesquisa relevante tenha sido concluída. Como Sober coloca, “O único profilático de que precisamos é o empirismo” (SOBER, 2005, 97).

No entanto, Sober também argumenta que a metodologia empírica da psicologia coloca o ônus da prova em maior grau na pesquisa da cognição animal do que na pesquisa da cognição humana. Ele sugere que os psicólogos comparativos aceitam como hipótese nula que diferentes mecanismos cognitivos estão em ação em humanos e animais. Dado que os erros do tipo 1, relatar um falso positivo e rejeitar uma hipótese nula (possivelmente verdadeira), são considerados mais graves do que os erros do tipo 2 (relatar um falso negativo e não rejeitar uma hipótese nula quando ela é falsa), a prática da ciência resulta em um enviesamento contra a atribuição de traços psicológicos aos animais (SOBER, 2005).

O debate sobre como interpretar os resultados dos estudos com animais em comparação com os estudos com humanos pode ser visto como um debate sobre uma aplicação inconsistente do que o psicólogo C. Lloyd Morgan propôs como seu Cânone. O Cânone de Morgan afirma: “Em nenhum caso podemos interpretar uma ação como o resultado do exercício de uma faculdade psíquica superior, se ela pode ser interpretada como o resultado do exercício de uma que

está mais abaixo na escala psicológica” (MORGAN, 1894, 53). Embora esta seja uma regra prática tradicional na pesquisa da cognição animal, às vezes referida como o “princípio de conservadorismo”, não é um princípio comumente usado na pesquisa da cognição humana. Para complicar as coisas, as tentativas de determinar quais faculdades psíquicas são superiores ou inferiores, uma tarefa que o Cânone de Morgan instrui pesquisadores a realizar, levantaram preocupações sobre sua significância e utilidade (ALLEN-HERMANSON, 2005; FITZPATRICK, 2008; SOBER, 2005).

Alguns argumentam que o antropomorfismo é uma tendência humana que deve ser superada para fazermos boa ciência, porque ele depende de uma generalização injustificada de humanos linguísticos para animais não linguísticos. Esses críticos sugerem que os animais sem linguagem não podem sequer ter conceitos, e sem a linguagem os cientistas não estão em posição de atribuir conteúdo. Dado que estamos barrados de fazer tais atribuições, a psicologia científica não deve se envolver com questões sobre a mentalidade animal (*vide* KEETON, 1967; KENNEDY, 1992; BLUMBERG; WASSERMAN, 1995; WYNNE, 2004, 2007).

Foi notado que tais argumentos dizem respeito aos métodos apropriados para a ciência, ao escopo da ciência e a como interpretar os dados (KEELEY, 2004; BEKOFF; ALLEN, 1997). Sendo assim, tais argumentos não são empíricos, mas teóricos ou metodológicos. Isso pode ser visto na maneira como os debates às vezes resultam em um impasse. Aqueles que se opõem a atribuir propriedades mentais aos animais são acusados de petição de princípio (GRIFFIN, 1992), por cometer “antropocentrismo reverso” (SHEETS-JOHNSTONE, 1992), “antroponegação” (DE WAAL, 1999) ou “antropsectomia” (ANDREWS; HUSS, 2014).

Essas preocupações surgem na formulação da hipótese nula em pesquisas experimentais. Devido aos métodos estatísticos padrão usados em experimentos de cognição animal, a escolha da hipótese nula em termos de um animal sem capacidade cognitiva envia a pesquisa contra a descoberta de evidências a favor dessa capacidade (ANDREWS; HUSS, 2014; MIKHALEVICH, 2014; SOBER, 2005). Uma preocupação é que os pesquisadores podem ter uma falha de imaginação quando se trata da geração de hipóteses; eles podem fazer um argumento de inferência pela melhor explicação sem considerar um número suficiente de explicações plausíveis. Isso reflete a preocupação de Kennedy quando ele afirma que o seguinte argumento para atribuir propriedades mentais a animais repousa em uma falsa dicotomia: ou os animais são máquinas de estímulo-resposta ou são agentes com crenças e desejos; dado que os animais não são máquinas de estímulo-resposta, eles devem ser agentes psicológicos (KENNEDY, 1992). De acordo com Kennedy,

o problema com esse argumento é que nem todas as máquinas implementam funções de estímulo-resposta; algumas máquinas são complexas e indeterministas, e se os animais fossem máquinas, seriam máquinas desse tipo (BARLOW, 1990; KENNEDY, 1992). Preocupações similares são apresentadas por aqueles que enfatizam, contra Darwin, a descontinuidade entre humanos e outros animais (PENN *et al.*, 2008).

Finalmente, alguns temem que a ciência da cognição comparativa tenha sido prejudicada pela ideia de que a explicação do comportamento em termos de predisposições típicas da espécie, em vez de em termos de raciocínio ou *insight* humano, equivale a uma explicação “estraga-prazeres” (SHETTLEWORTH, 2010b). Quando se considera as diferenças entre humanos e animais como menos satisfatórias do que as semelhanças, não deveria ser surpresa se mais semelhanças forem encontradas. Pior ainda é a preocupação de que algum “raciocínio semelhante ao humano” postulado para explicar o comportamento humano seja falso e que comportamentos extraordinários realizados por humanos e outros animais possam surgir de processos simples (ANDREWS, 2005, 2012; BARRETT, 2011; BUCKNER, 2013; PENN *et al.*, 2008; SHETTLEWORTH, 2010b; de WAAL; FERRARI, 2010). A ideia de que as capacidades cognitivas humanas são frequentemente exageradas e “superintelectualizadas” é inspirada por trabalhos em vários campos, incluindo abordagens em ciências cognitivas que se concentram no poder de regras simples e no surgimento de comportamentos complexos em sistemas auto-organizadores, propriedades emergentes, representações mentais, vida artificial, robótica, assim como abordagens situadas, distribuídas e dinâmicas para a cognição, a literatura de heurísticas e vieses em cognição social, além de teorias de processo dual da arquitetura cognitiva. Buckner chama o erro de comparar as capacidades cognitivas dos animais com uma imagem exagerada das capacidades cognitivas humanas de “antropofabulação”, uma vez que envolve dois erros somados: confabular mecanismos cognitivos humanos mais antropocentrismo (BUCKNER, 2013).

Em geral, as preocupações com o antropomorfismo parecem ser limitadas aos cientistas ocidentais. Tem-se argumentado que pesquisadores de países com orientação budista, em vez de cristã, não são culturalmente encorajados a ver uma distinção categórica entre humanos e animais não humanos (ASQUITH, 1991; SAKURA, 1998; MATSUZAWA, 2003; DE WAAL, 2003). Diferente do cristianismo, a doutrina budista não afirma que os humanos (mas não os animais) têm almas imortais e, portanto, não permite que os humanos usem os animais para seus próprios fins da maneira como o cristianismo permite. A tradição budista vê uma conexão entre humanos e outros animais, e até afirma que humanos podem renascer

como animais. De Waal argumenta que a diferença nas atitudes culturais em relação aos animais levou os primatologistas ocidentais a inicialmente rejeitar os métodos, descobertas e ideias japonesas; só recentemente algumas dessas ideias, como a alegação de Kinji Imanishi de que os primatas exibem diferenças culturais dentro das espécies, conquistaram seu espaço no discurso científico ocidental (DE WAAL, 2003).

## 2.3 Racionalidade

Ações racionais são ações que de alguma forma se encaixam entre si. Embora muitas vezes pensada como o resultado de um raciocínio lógico, a ação racional também pode ser entendida como teleológica, causal ou probabilística. Para uma coletânea de ensaios sobre racionalidade em diferentes espécies, ver Hurley & Nudds (2006).

O filósofo estoico Crísipo sugeriu que podemos ver o raciocínio lógico no comportamento animal, através de sua história do cachorro que, correndo com o nariz no chão, rastreou um coelho por um caminho. Quando o cão chega a uma encruzilhada de três vias, ele rapidamente fareja os dois primeiros caminhos e, não encontrando o cheiro em nenhuma das duas primeiras opções, corre imediatamente pelo terceiro caminho, sem cheirar primeiro. Essa história sugere que o cachorro de Crísipo foi capaz de fazer uma inferência racional da forma:

- 1 A ou B ou C.
- 2 Não A.
- 3 Não B.
- 4 Logo, C.

No entanto, o cão pode ter resolvido o problema sem um raciocínio lógico inteiramente desenvolvido. Teorias mínimas da racionalidade oferecem alternativas.

Algumas teorias mínimas da racionalidade derivam do pensamento evolucionário. Por exemplo, na visão de Fred Dretske, mesmo alguns comportamentos aprendidos simples, como um pássaro evitar comer uma borboleta monarca, podem ser interpretados como minimamente racionais. Como as monarcas que comem erva-leiteira tóxica se tornam tóxicas para pássaros e outros predadores, quando um pássaro aprende a não comer borboletas monarca depois de formar uma associação entre comer monarcas e vomitar, há um motivo para seu comportamento



de evitação. Os pássaros também têm um motivo para evitar comer uma vice-rei, visto que é visualmente quase indistinguível de uma monarca, embora não seja venenosa. O comportamento em ambos os casos é explicado pelo conteúdo do pensamento (ou "pensamento") do pássaro, e para Dretske isso é suficiente para o pássaro contar como um agente minimamente racional (1988, 2006). Outras teorias da racionalidade que levam em conta considerações evolutivas incluem as de Daniel Dennett (1995, 1987), Ruth Millikan (2004, 2006) e Joelle Proust (1999, 2006).

As teorias causais da ação racional animal retratam o animal como engajado em um raciocínio causal, em vez de proposicional. José Bermúdez argumenta que nem toda ação animal pode ser explicada meramente em termos de racionalidade mínima ou de teleologia (2003, 2006). Em vez disso, existem alguns comportamentos que só podem ser explicados em termos de atitudes proposicionais, estados informacionais ou generalizações que vão além do "aqui e agora". No entanto, uma vez que os animais não podem se envolver em metacognição e pensar pensamentos sobre pensamentos, eles não podem ter os conceitos de inferência necessários para o raciocínio lógico. Em vez disso, podemos descrever seu processo de raciocínio em termos causais. Considere uma gazela que vê um leão e depois foge. As gazelas podem entender que os leões causam que elas corram, e que, como há um leão aqui, eu corro. Essa compreensão causal é desenvolvida por meio da experiência com regularidades no ambiente, e embora elas sejam capazes de generalizar em alguma medida, essa capacidade é limitada.

Teorias probabilísticas da racionalidade animal também são usadas para explicar o comportamento animal complexo. Por exemplo, como uma explicação do caso do cachorro de Crísipo, Michael Rescorla sugere que o cachorro poderia ter empregado raciocínio Bayesiano, podendo formar e atualizar probabilidades perante mudanças nas informações perceptivas (RESCORLA, 2009). Modelos de cognição por codificação preditiva que não dependem do modelo de processamento linguístico associado à teoria computacional da mente podem ser usados para explicar o comportamento animal como sendo racional no mesmo patamar da racionalidade humana.

A racionalidade em outras espécies foi explorada em estudos experimentais e naturalísticos. Psicólogos testaram formalmente animais em situações como a do cachorro de Crísipo. Por exemplo, os grandes símios parecem engajar-se no raciocínio de exclusão: quando sabem que um de dois recipientes opacos estão iscados com comida, ao serem mostrados que o vazio está vazio, vão imediatamente pegar o segundo recipiente, sem olhar dentro dele (CHAMAR, 2004, 2006; MARSH

E MACDONALD, 2011; ERDŐHEGYI *et al.*, 2007). Há evidências de que macacos, corvídeos e cães também podem, em alguns casos, escolher por exclusão.

Certos comportamentos naturalísticos também sugerem pensamento racional, visto que parecem ser casos de resolução de problemas que dependem de flexibilidade cognitiva e aprendizado. O uso de ferramentas, por exemplo, é um comportamento que sugere pensamento racional. Como o uso da ferramenta envolve encontrar ou construir um objeto que é utilizado como uma extensão do corpo para atingir um objetivo específico, o uso da ferramenta envolve a identificação de um problema, a consideração de formas de resolver o problema, e a percepção de que outros objetos podem ser usados na manipulação da situação. As primeiras pesquisas experimentais sobre resolução de problemas por chimpanzés, feitas pelo psicólogo alemão Wolfgang Köhler, tinham chimpanzés construindo ferramentas para adquirir objetos fora de alcance. Foi relatado que os chimpanzés empilhavam caixas ou montariam tubos para formar uma longa haste a fim de alcançar as bananas penduradas no alto (KÖHLER, 1925). Dado esse comportamento, Köhler sugeriu que os chimpanzés resolvem alguns problemas não por tentativa e erro ou por associação estímulo-resposta, mas por meio de um *insight* repentino. Confirma Povinelli (2000) para uma crítica da interpretação contemporânea da pesquisa de Köhler.

Os estudos naturalísticos sobre o uso de ferramentas por animais decolaram na década de 1960, quando duas equipes de pesquisa independentes na Tanzânia observaram chimpanzés fazendo e usando ferramentas. Goodall encontrou chimpanzés em Gombe usando gramineas e galhos para pescar cupins, e observou os chimpanzés modificando os galhos e arrancando suas folhas para que pudessem ser usados para esse propósito (GOODALL, 1986). Na mesma época, a equipe de Kinji Imanishi encontrou chimpanzés usando pedras para quebrar nozes na floresta de Mahale, a cerca de 320 quilômetros do local de pesquisa de Goodall (NISHIDA, 1990).

Agora sabemos que os chimpanzés fabricam e usam ferramentas para diversos fins. Os chimpanzés em Fongoli (Senegal) fabricam lanças em quatro ou mais etapas para caçar galagos (PREUTZ; BERTOLANI, 2007). Chimpanzés em Bossou (Guiné) usam grandes galhos de árvores dendezeiro para abrir a árvore em sua copa, a fim de obter acesso a uma rica fonte de alimento (YAMAKOSHI; SUGIYAMA, 1995). Os chimpanzés também constroem e usam conjuntos de ferramentas que posteriormente utilizam em uma ordem determinada: os chimpanzés de Goualougo fabricam uma ferramenta de perfuração para aumentar os buracos de um ninho de cupins após uma tentativa de pesca malsucedida; assim que o

buraco de saída é alargado, o chimpanzé insere uma sonda de pesca (SANZ; MORGAN, 2007). Os chimpanzés também usam compostos de ferramentas, como martelo e bigorna, para quebrar nozes (NISHIDA, 1990; SAKURA; MATSUZAWA, 1991) e fabricam ferramentas de pedra (CARVALHO *et al.*, 2008).

O uso de ferramentas na natureza foi descoberto em diferentes táxons, incluindo invertebrados como polvos, pássaros, peixes, anfíbios, répteis, mamíferos não primatas, macacos e grandes símios (SHUMAKER *et al.*, 2011). Relatos de uso de ferramentas com animais oferecem evidências a favor da alegação de que algum comportamento animal é funcionalmente racional, no sentido de que seu comportamento permite que o animal atinja um objetivo. Além disso, talvez seja o resultado de uma adaptação evolutiva. No entanto, a extensão em que tais evidências abordam a questão de saber se o comportamento é racional vai depender da visão de alguém sobre a natureza da racionalidade.

## 2.4 Crença

A questão de saber se os animais têm crenças também é difícil de responder devido à falta de acordo teórico sobre a natureza da crença. A visão mais comum é a de que a crença é um estado representacional, e que a representação mental, que fixa o conteúdo, expressa um conteúdo proposicional. Para alguns, essa visão é consistente com a crença animal, uma vez que acreditam que, como os humanos, os animais podem operar em uma linguagem de pensamento (FODOR, 1975; CHENEY; SEYFARTH, 2007; TETZLAFF; RAY, 2009). Alguns representacionistas sugeriram que os animais podem ter crenças representacionais que não podem ser expressas proposicionalmente (BERMÚDEZ, 2003; CAMP, 2009; RESCORLA, 2009). Por exemplo, Camp sugere que algumas crenças são baseadas em sistemas representacionais imagéticos, como diagramas ou mapas, e que tais sistemas podem explicar o conhecimento social dos babuínos, e então não há necessidade de postular uma linguagem de pensamento babuíno. Confira Camp (2009) em resposta para Cheney e Seyfarth (2007). Enquanto Cheney e Seyfarth afirmam que o sistema cognitivo do babuíno é veiculado por uma linguagem de pensamento dada sua natureza como representacional, de valor discreto, hierarquicamente estruturado, governado por regras, aberto, e independente da modalidade sensorial, Camp argumenta que a inferência não é justificada, pois sistemas representacionais não frasais também podem ter tais recursos. Babuínos podem ter crenças sem

terem crenças com conteúdo proposicional. Bermúdez também argumenta que os animais têm crenças não proposicionais, porque alguns comportamentos animais só são explicáveis em termos de crenças e desejos. No entanto, Bermúdez nega que a crença animal tenha qualquer estrutura; em vez disso, em sua proposta, crenças animais são simples representações imagéticas de coisas no mundo que permitem pensamentos sobre os objetivos e percepções dos outros (BERMÚDEZ, 2003). A visão de que os animais podem ter crenças perceptivas que não devem ser entendidas em termos de atitudes proposicionais também foi defendida por Glock (2000, 2010).

Representacionistas de toda sorte apontam as limitações dos diferentes tipos de sistemas representacionais. Camp, que defende que os mapas têm uma estrutura sintática rica, admite algumas limitações: mapas mentais não parecem ser capazes de acomodar alguns pensamentos de primeira ordem, tais como proposições não-específicas quantificadas existencialmente ou como proposições quantificadas universalmente, e nem pensamentos de segunda ordem (CAMP, 2007). Como Camp, Bermúdez defende a posição de que, embora os animais possam pensar, eles não podem ter crenças sobre crenças; mas ele nega que o pensamento imagético tenha estrutura sintática e, portanto, conclui que os animais não são pensadores lógicos (BERMÚDEZ, 2003, 2009). Carruthers também argumenta que, embora até mesmo os insetos tenham crenças (CARRUTHERS, 2004), não há evidências de que qualquer espécie não humana tenha habilidades metacognitivas (CARRUTHERS, 2008). Carruthers sugere que a evidência oferecida pelo comportamento animal sustenta a posição de que a crença, como o desejo, vem em graus (CARRUTHERS, 2008).

Para outros, a falta de uma linguagem externa (e a falta de uma teoria do erro para explicar por que os animais teriam uma linguagem de pensamento enquanto carecem de linguagem externa) sugere que os animais não têm nenhuma crença (DENNETT, 1996; DAVIDSON, 1982; STICH, 1978). Donald Davidson ofereceu um argumento contra a racionalidade animal com base em uma associação entre conceitos, crenças e linguagem. Na visão de Davidson, seres que têm crença devem ter o conceito de crença, porque para se ter uma crença deve-se reconhecer que crenças podem ser verdadeiras ou falsas, e não se pode compreender a verdade objetiva sem se compreender a natureza de crenças. Para desenvolver uma compreensão da verdade objetiva, deve-se ser capaz de triangular com outros, de falar com outros sobre o mundo e, portanto, todos os seres que têm crença devem ser usuários de uma linguagem. Visto que outras espécies carecem de linguagem,

elas não têm crenças (DAVIDSON, 1982). Davidson também argumenta contra as crenças animais com base na alegação de que, para alguém contar como um ser que tem crença, é necessário possuir alguma noção de erro (DAVIDSON, 1975). Esses argumentos foram amplamente discutidos na literatura.

Um argumento diferente contra crenças animais foi apresentado por Stephen Stich, que argumenta que não podemos atribuir atitudes proposicionais aos animais em qualquer sentido metafisicamente robusto, dada a nossa incapacidade de atribuir conteúdo à suposta crença de um animal (STICH, 1978). Na visão de Stich, se a atribuição de crenças aos animais é entendida de modo puramente instrumental, então os animais têm crenças. No entanto, se a atribuição de crenças aos animais requer que possamos descrever com precisão o conteúdo dessas crenças, então os animais não têm crenças. Em resposta à sugestão de Armstrong, de que podemos fixar o conteúdo *de re* da crença animal (ARMSTRONG, 1973), Stich argumenta que não podemos fazer atribuições *de re* porque isso violaria o papel de preservação da verdade da atribuição. Além disso, dado que "nada que pudéssemos descobrir nos permitiria atribuir conteúdo à crença de um animal" (STICH, 1978, p. 23), não podemos fazer atribuições *de dicto* a outras espécies. Portanto, não podemos fazer nenhuma atribuição, e se não podemos dizer do que se trata a crença de um animal, não faz sentido dizer que um animal tem uma crença. A preocupação aqui é similar à preocupação com o antropomorfismo; quando usamos nossa linguagem para atribuir conteúdo a outras espécies, podemos estar atribuindo a elas mais do que o apropriado. Stich se preocupa com o fato de que, quando dizemos "Fido acredita que há um osso carnudo enterrado no quintal", estamos atribuindo a Fido conceitos que ele não pode ter; conceitos como **quintal**, que só são compreensíveis se tivermos conceitos correspondentes como **linhas de divisa**, **casa**, **cerca**, e assim por diante. O argumento de Stich pode ser formulado como:

- 1 Para que algo tenha uma crença, deve ter um conceito.
- 2 Para se ter um conceito, é necessário ter certos tipos específicos de conhecimento, incluindo o conhecimento de como o conceito se relaciona com outros conceitos.
- 3 Animais não humanos não têm esse conhecimento.
- 4 Portanto, os animais não humanos não têm crenças.

Existem pelo menos três maneiras de responder aos argumentos de que as crenças requerem conceitos que requerem linguagem. Pode-se negar a conexão

necessária entre conceitos e linguagem, como vimos acima; também, como veremos na próxima seção, alguns argumentam que conceitos podem ser possuídos por indivíduos que não têm linguagem (ALLEN; HAUSER, 1996; ALLEN, 1999; GLOCK, 2010). A segunda maneira é negar a conexão necessária entre crenças e conceitos; dado o conteúdo não conceitual, pode ser possível crer sem ter conceitos (BERMÚDEZ, 2003; GLOCK, 2010). Um terceiro caminho é ser um instrumentalista sobre crenças, tal que ser capaz de atribuir conteúdo de modo a prever comportamentos corretamente é suficiente para haver crença, mesmo na ausência de linguagem (DENNETT, 1995), ou atribuindo um conteúdo proposicional que rastreie o conteúdo das mentes dos animais (ROWLANDS, 2012). Pode-se chegar ao ponto de argumentar que crenças não precisam ser entendidas em termos de representações mentais. Além disso, dado que algumas teorias de determinação de conteúdo não dependem de habilidades linguísticas (por exemplo, teleológicas, causais), pode-se tentar contornar os tipos de preocupações levantadas por Davidson e Stich adotando tais teorias.

## 2.5 Conceitos

Enquanto alguns filósofos defendem a afirmação metafísica de que a linguagem natural é necessária para possuir conceitos (BRANDON, 1994; DAVIDSON, 1975; DUMMETT, 1993), e enquanto outros defendem a afirmação epistêmica de que a posse da linguagem é uma condição necessária para identificar conceitos (STICH, 1978), pesquisas empíricas em cognição animal sugerem que ambas as visões podem estar erradas. Os pesquisadores buscam descobrir a natureza dos conceitos animais tanto *de re* como *de dicto* usando uma variedade de métodos.

Um dos primeiros métodos para examinar os conceitos de animais surgiu de uma série de experimentos com pombos. Fotografias foram mostradas aos pombos, e eles foram treinados para bicar as imagens que continham um dado objeto, como uma árvore, e não responder às imagens que não continham o dado objeto. Após o período de treinamento, os pombos puderam generalizar para novas fotografias, bicando apenas aquelas que continham árvores exatamente como no treinamento. Foi sugerido que essa habilidade de classificação demonstra que o pombo tem um conceito do objeto alvo (HERRNSTEIN, 1979; HERRNSTEIN *et al.*, 1976).

Muitos rejeitam a ideia de que ser capaz de organizar objetos é suficiente para possuir um conceito correspondente ao categorial, uma vez que os humanos podem organizar objetos sem possuir o conceito correspondente. Como Colin Allen

e Marc Hauser colocam: “É possível ensinar um ser humano a separar entre distribuidores e outras peças de motores de automóvel, com base na semelhança de família entre os formatos dos distribuidores. Mas essa capacidade não seria suficiente para dizermos que a pessoa possui o conceito de distribuidor” (ALLEN; HAUSER, 1996, 51).

Em vez de identificar a posse de conceitos com comportamentos de organização, Allen e Hauser sugerem metodologias alternativas para identificar conceitos em outras espécies. Por exemplo, eles oferecem um possível (embora, eles admitem, eticamente insustentável) teste para um conceito de morte entre macacos-vervet (ALLEN; HAUSER, 1996). As mães de macacos-vervet são capazes de reconhecer os guinchos de alerta de seus bebês e, quando ouvem esse choro, a mãe olha para o filho e as outras fêmeas olharão para a mãe. Allen e Hauser sugerem que reproduzir uma gravação do guincho de alerta de um bebê recentemente falecido ajudaria a determinar se os macacos-vervet têm um conceito de morte. Se a mãe responde a essas gravações de maneira atípica, ao contrário da resposta usual feita a um bebê vivo, essa resposta fornece evidências de que os macacos-vervet têm o conceito de morte. De acordo com Allen e Hauser, ter um conceito permite respostas diferentes a estímulos idênticos. O som do guincho de alerta do bebê durante a vida é idêntico ao som reproduzido após a morte. Se a resposta a esse estímulo for diferente, isso é evidência de que houve uma mudança conceitual associada ao estímulo. Allen apresenta a estratégia geral para atribuir conceitos aos animais da seguinte forma: "Um conceito de X (por exemplo, ÁRVORE) pode ser razoavelmente atribuído a um organismo O sempre que:

- i O discrimina sistematicamente alguns Xs de alguns não-Xs; e
- ii O é capaz de detectar alguns de seus próprios erros de discriminação entre Xs e não Xs; e
- iii O é capaz de aprender a discriminar melhor os Xs dos não-Xs como consequência de sua capacidade" (ALLEN, 1999, 37).

Outro método para estudar conceitos de animais vem de pesquisas em bebês humanos (HAUSER *et al.*, 1996; HAUSER; CAREY, 1998; BERMÚDEZ, 2003; GÓMEZ, 2005; ULLER, 2004). O paradigma do olhar preferencial, também conhecido como paradigma da desabituação, é usado para estudar a compreensão

que bebês humanos têm do mundo físico e social (BAILLARGEON; DEVOS, 1991; SPELKE, 1991). Acredita-se que os experimentos de desabitação nos ajudem a entender que tipos de predições os bebês fazem sobre o mundo, o que pode nos ajudar a compreender como eles veem o mundo. A metodologia é simples: um estímulo é repetidamente mostrado a um bebê e, assim que ele se habitua ao estímulo, o bebê se torna desinteressado. Neste momento, um novo estímulo é mostrado. Se o bebê vê o novo estímulo como diferente do estímulo inicial, ou como impossível dado o estímulo inicial, ele vai olhar durante mais tempo para o novo estímulo. Se o bebê considerar o novo estímulo como semelhante ao estímulo inicial, ele não mostrará nenhum interesse adicional. A ideia é que, comparando as respostas entre grupos de indivíduos, um pesquisador pode aprender algo sobre como aquele grupo conceitua o mundo. Bermúdez argumenta que tais métodos podem ser usados para fazer atribuições *de dicto* a animais (BERMÚDEZ, 2003b).

Em um estudo usando este método, Marc Hauser e colegas investigaram conceitos numéricos em diferentes espécies de primatas, incluindo macacos-rhesus (HAUSER *et al.*, 1996) e saguins-cabeça-de-algodão (ULLER, 1997). Os pesquisadores testaram a capacidade dos macacos de rastrear objetos individuais colocados atrás de uma barreira. Eles descobriram que, como bebês humanos, os macacos olham mais longamente para resultados impossíveis. Por exemplo, em uma condição de teste, os macacos-rhesus viram duas berinjelas sendo colocadas sequencialmente atrás de uma tela e, em seguida, a tela foi levantada mostrando apenas uma berinjela. Os macacos olharam por mais tempo para a berinjela individual do que quando foram mostradas duas berinjelas, sugerindo que eles representam as berinjelas como categoriais distintos.

Outra maneira de aprender como diferentes espécies organizam o mundo é ensinando aos indivíduos um sistema de comunicação simbólica. Por exemplo, a bióloga Irene Pepperberg treinou um papagaio cinza africano chamado Alex para classificar objetos usando conceitos de metanível que categorizam outros conceitos. Alex foi capaz de classificar objetos de acordo com a cor, forma e matéria, e foi capaz de classificar conjuntos de objetos de acordo com o número. Além de classificar, Alex poderia relatar qual característica torna dois objetos semelhantes ou diferentes. Por exemplo, quando apresentado a um bloco vermelho e uma chave vermelha, Alex respondeu à pergunta "O que é igual?" proferindo "cor". Ele também pôde relatar semelhanças e diferenças de forma e matéria. Pepperberg afirma que seus estudos demonstram que Alex compreende conceitos categóricos e revelam as classificações que Alex inventou (PEPPERBERG, 1999). Entretanto, uma preocupação



é a de que os conceitos exibidos por animais treinados com símbolos sejam um artefato do sistema de comunicação, em vez de típicos da espécie.

Finalmente, os conceitos de animais também estão sendo investigados em estudos de campo, onde a utilidade dos conceitos para os animais pode ser mais aparente. Por meio de observação cuidadosa e experimentos de campo com animais selvagens, pesquisadores examinam os conceitos naturais que diferentes espécies podem usar para categorizar eventos em seu ambiente social e físico. Por exemplo, primatologistas observaram que os macacos-de-Diana possuem chamados de alerta acusticamente distintos para predadores diferentes, e que os chamados de alerta femininos diferem dos masculinos (ZUBERBÜHLER *et al.*, 1997, 1999). Essa diferença natural permitiu que Zuberbühler e colegas montassem um experimento de campo com o objetivo de determinar se os guinchos de alerta masculino e feminino significam a mesma coisa para os macacos. Ele descobriu que os macacos tratavam os chamados acusticamente diferentes como sinônimos. Tal evidência sugere que os pesquisadores podem identificar pelo menos alguns conceitos "*de re*" de animais.

O conhecimento social oferece uma janela para pesquisadores de campo que investigam conceitos animais. Cheney e Seyfarth argumentam que o comportamento dos primatas depende de um rico corpo de conhecimento social, e tal conhecimento sugere que os primatas possuem compreensão conceitual (CHENEY; SEYFARTH, 2009, 2015). No caso dos babuínos, sabemos que eles reconhecem indivíduos, classificando-os em grupos de acordo com suas propriedades, incluindo laços sociais íntimos, parentesco, hierarquia de dominância e relações sexuais transitórias. Por exemplo, o conhecimento de parentesco é demonstrado em casos de reconciliação mediada por parentes, quando um encontro antagônico é resolvido por um parente do agressor dando um grunhido de reconciliação. Esta informação categórica informa o comportamento babuíno. Além disso, alguns desses relacionamentos mudam com o tempo, e podem ter repercussões difundidas na organização de dominância do grupo. Os babuínos são capazes de fazer rápidos ajustes nas classificações de dominância lineares após uma reversão de hierarquia, mesmo quando a reversão afeta matrilineas diferentes e causa mudanças na relação de classificação de vários indivíduos. Cheney e Seyfarth argumentam que a memória e o condicionamento clássico por si só não podem explicar a riqueza do conhecimento social dos primatas, dada a quantidade de informações que os primatas teriam de representar. Eles afirmam que um babuíno teria que aprender milhares de relações diádicas e dezenas de milhares de relações triádicas a fim de antecipar o comportamento de outros animais. Além da preocupação com o espaço necessário para representar todas

essas relações, eles apontam que a velocidade do comportamento dos babuínos em resposta a um problema complexo não é consistente com a hipótese de que os babuínos resolvem problemas sociais pesquisando em um banco de dados enorme e não estruturado de relações. Em vez disso, Cheney e Seyfarth sugerem que babuínos e outros primatas com sociedades sociais complexas organizam os indivíduos em classes ou conceitos governados por regras. Isso, eles argumentam, é uma estratégia adaptativa (CHENEY; SEYFARTH, 2009).

Embora grande parte da discussão tenha se concentrado em se os animais têm conceitos e em como identificar tais conceitos, também foi argumentado que a evidência animal sugere que podemos representar melhor a diversidade cognitiva e a complexidade individual se considerarmos que existem diferentes tipos de representações, e nem todas elas são conceituais. Alguns comportamentos animais podem depender de conteúdo não conceitual que permite um comportamento flexível, mas não requer a habilidade de identificar objetos, enquanto outros comportamentos dependem de conceitos baseados na percepção e até mesmo de conteúdo proposicional, por exemplo, no caso de macacos treinados por símbolos (NEWEN; BARTELS, 2007).

A ideia de que, se não podemos dizer o que os animais pensam, então os animais não têm crenças, foi questionada por apelo a conteúdo não conceitual. Jacob Beck (2013) sugere que não podemos dizer o que um animal pensa já que os animais pensam em um formato não linguístico. Assim como uma pintura pode ter conteúdo sem ter conteúdo expressável na linguagem, os animais podem ter conteúdo não expressável na linguagem. Além disso, os humanos podem compartilhar parte desse conteúdo com animais e, portanto, poderíamos compartilhar crenças com um animal se pensarmos sobre a mesma coisa no mesmo formato não linguístico. O formato não linguístico pode ser analógico, em vez de um formato digital, como o idioma. Os formatos analógicos não permitem a análise em termos de como as partes significativas se encaixam para formar um todo significativo, como a linguagem faz. Em vez disso, os formatos analógicos têm significado holístico, mas, como uma fotografia, as representações analógicas vêm em graus de foco. Exemplos de conteúdo analógico incluem fotos, imagens e mapas, e como não podemos traduzir essas representações para uma frase (Beck nos pede para considerar como traduzir a Mona Lisa para o inglês!), não somos capazes de traduzir representações de animais para a linguagem. Mas os animais ainda podem ter representações, assim como a Mona Lisa ainda existe, apesar de ser intraduzível. Beck fornece um exemplo de conteúdo não conceitual compartilhado por humanos e animais no caso de

estados de magnitude analógicos, que são representações de números aproximados (BECK, 2012). Os humanos e outras espécies podem fazer julgamentos sobre arranjos maiores ou menores de objetos, desde que a proporção seja grande o suficiente. Beck argumenta que essas representações de números aproximados não podem ser expressas com precisão em sentenças e conclui que isso oferece evidências a favor de conteúdo não conceitual para animais e humanos. Magnitudes analógicas são exemplos de estados representacionais perceptivos, que não têm as mesmas propriedades lógicas dos estados representacionais sentenciais.

### **3. A Ciência da Cognição Animal**

Atualmente, nenhuma disciplina detém o monopólio do estudo da cognição animal. Psicologia, biologia, antropologia, bem-estar animal, filosofia, estudos de animais e outros programas incluem pesquisadores que trabalham com cognição animal. Assim como nenhuma disciplina tem um foco exclusivo na cognição animal, não existe um método que caracterize a pesquisa em cognição animal. Enquanto o interesse científico inicial no comportamento animal foi abordado de modo diferente por psicólogos, que se concentraram no uso de experimentos de laboratório para descobrir mecanismos, e por etologistas, que usaram métodos observacionais para descobrir a evolução do comportamento, hoje muitos cientistas abordam o estudo da cognição animal usando uma combinação de métodos.

#### **3.1 Método Anedótico Inicial**

O interesse científico contemporâneo em mentes animais e capacidades cognitivas frequentemente remonta à teoria da evolução por seleção natural de Charles Darwin. Em *The Descent of Man*, Darwin introduziu muitos dos tópicos que motivam os programas de pesquisa em cognição animal atualmente, incluindo o uso de ferramentas, raciocínio, aprendizagem, conceitos, consciência, cognição social, habilidades artísticas e cognição moral. Além disso, Darwin antecipou o interesse atual no raciocínio implícito com seu comentário “O selvagem certamente não saberia e nem se importaria por qual lei os movimentos desejados foram efetuados; ainda assim, seu ato seria guiado por um processo rudimentar de raciocínio, tão certo quanto o de um filósofo em sua mais longa cadeia de deduções” (DARWIN, 1974, p. 75).

Como Aristóteles séculos antes dele, Darwin defendeu a continuidade do mental entre as espécies; assim como algumas características morfológicas são homólogas entre espécies que vivem em ambientes semelhantes, é de se esperar que também haja semelhanças psicológicas e comportamentais: "a diferença de mente entre o homem e os animais superiores, por maior que seja, certamente é de grau e não de tipo" (DARWIN, 1974, p. 122). Esta visão também foi defendida pelo contemporâneo de Darwin, o naturalista George Romanes, que em seu livro *Animal Intelligence* escreve "deve haver uma continuidade psicológica, assim como fisiológica, que se estende por todo a profundidade e amplitude do reino animal" (ROMANES, 1970, p. 10). A visão de Darwin sobre a continuidade entre as espécies foi desafiada em várias dimensões, desde a ideia de que a língua ou cultura desempenha um papel fundamental na evolução da cognição, até as preocupações com o antropomorfismo (ANDREWS; RADENOVIC, 2012; PENN *et al.*, 2008; SHETTLEWORTH, 2010b).

O método que Darwin, Romanes e seus contemporâneos usaram pela primeira vez para investigar essas questões pode ser descrito como o método anedótico. Histórias sobre o comportamento animal foram coletadas de uma variedade de pessoas, incluindo oficiais militares, naturalistas amadores e pessoas em geral. Tais histórias foram compiladas e usadas como evidência de uma capacidade cognitiva particular em dada espécie. Essa abordagem foi amplamente criticada. A "evidência" reunida era frequentemente uma história contada sobre um evento testemunhado por uma única pessoa, geralmente um observador sem treinamento científico. Além disso, essas histórias costumavam ser adquiridas de segunda ou terceira mão, e por isso havia a preocupação de que os relatórios tivessem sido embelezados ou alterados de alguma forma ao longo do caminho. Esses problemas foram reconhecidos logo de início e, em resposta, Romanes desenvolveu três princípios para aceitar anedotas, a fim de evitar alguns desses problemas:

- i Nunca aceite um relato de incidente como fato sem considerar a autoridade ou respeitabilidade do observador.
- i Se o observador não for conhecido e o relatório do incidente for suficientemente importante, considere se o observador pode ter razão ou causa para fazer um relatório errôneo.

- iii Procure corroborações da observação examinando observações semelhantes ou análogas feitas por outros observadores independentes (ROMANES, 1970).

O terceiro princípio foi aquele em que ele mais se baseou, escrevendo “Este princípio descobri ser de grande utilidade para guiar minha seleção de casos, pois, quando afirmações factuais que não apresentam nada de intrinsecamente improvável são independentemente confirmadas por diferentes observadores, elas têm o direito de serem consideradas confiáveis, tanto quanto declarações que se baseiam na autoridade única de um observador conhecido, e eu descobri que as primeiras são pelo menos tão abundantes quanto as últimas” (ROMANES, 1970, p. IX).

Apesar das tentativas de Romanes, o método permaneceu problemático, na medida em que não fornecia nenhuma informação estatística sobre a frequência de tais comportamentos; o viés de seleção levaria as pessoas a relatar apenas os comportamentos inteligentes interessantes e a ignorar a frequência dos comportamentos que podem servir como contra-evidências. Assim, o método anedótico praticado por Darwin e Romanes carece de muitas das virtudes associadas aos bons métodos científicos.

Embora o método anedótico inicial tenha sido rejeitado, alguns pesquisadores argumentam que as anedotas desempenham um papel importante na pesquisa da cognição animal (BATES; BYRNE, 2007), e a observação do comportamento ainda é considerada uma ferramenta de pesquisa valiosa, como veremos a seguir.

### 3.2 Métodos Experimentais

O biólogo e psicólogo britânico C. Lloyd Morgan, a quem muitas vezes é creditado o surgimento dos métodos contemporâneos de cognição animal, criticou os métodos de Romanes. Ele observou que os comportamentos animais que nos interessam podem ser causados de várias maneiras, o que torna a busca pelos mecanismos causadores do comportamento uma tarefa importante para um psicólogo animal. Morgan nos pediu que considerássemos o exemplo de Tony, um filhote de Fox-Terrier que conseguia abrir o portão de seu jardim e escapar para a estrada. Tony teve sucesso nessa tarefa colocando a cabeça sob o trinco do portão, levantando o trinco e esperando que o portão se abrisse. Morgan sugere que é possível interpretar o comportamento de Tony em termos de ele ter um objetivo e saber como alcançá-lo; mas, como existem outras interpretações, seria muito precipitado tirar

essa conclusão. Talvez, em vez disso, Tony estivesse respondendo às oportunidades visuais do portão, vendo a trava como elevável sem possuir o objetivo de sair do jardim. Uma terceira possibilidade é de que Tony use princípios gerais de raciocínio para abrir o portão, e aplique seu conhecimento geral a esta situação específica. É apenas essa terceira interpretação que Morgan categoriza como racional, dada sua visão de que o pensamento racional é o pensamento conceitual que permite uma análise por meio de princípios gerais.

Morgan conclui que o comportamento de Tony não deve ser interpretado como racional, porque o comportamento de Tony é interpretado satisfatoriamente sem exigir que ele forme algum conceito mental de como resolver a tarefa. Este caso serve como um exemplo do cânone de Morgan: “em nenhum caso uma atividade animal deve ser interpretada em termos de processos psicológicos superiores, se ela pode ser interpretada de forma justa em termos de processos que estão mais abaixo na escala de evolução e desenvolvimento psicológico” (MORGAN, 1903, p. 292). O Cânone de Morgan é um princípio epistêmico que aconselha a explicar um comportamento em termos da capacidade cognitiva mais baixa possível, e Morgan pensa que o raciocínio em termos de experiência sensorial é inferior, enquanto raciocinar conceitualmente em termos de princípios gerais é um processo psicológico superior.

Apesar de ser comum interpretar o Cânone de Morgan como uma regra comportamentalista, Morgan estava comprometido com a existência de mentes animais. Em sua autobiografia, ele escreveu: “[...] ao longo de toda a investigação, do início ao fim, meu interesse central foi psicológico, tal como entendo o significado desta palavra. Meu objetivo tem sido chegar à mente do passarinho, do cachorro etc., e enquadrar generalizações com relação à evolução mental” (MORGAN, 1930, p. 249). Além disso, Morgan, como Romanes, defendeu a atribuição de atividades mentais humanas aos animais usando o método de interpretação via introspecção, e observou a necessidade de interpretar o comportamento animal. Mas, ao mesmo tempo, ele nos alertou em ter cautela antes de julgar que comportamentos (humanos ou animais) que parecem ser inteligentes sejam realmente inteligentes. Morgan escreveu: “Para interpretar o comportamento animal, deve-se aprender também a ver a própria mentalidade em níveis de desenvolvimento muito mais baixos do que o nível superior de autoconsciência reflexiva. Isso não é fácil, e tem um quê de paradoxo” (MORGAN, 1930, p. 250). Este compromisso, que foi apelidado de Desafio de Morgan (ANDREWS, 2015), surge do reconhecimento de Morgan de que é difícil para os humanos seguir seu conselho de não intelectualizar demais a cognição humana. O erro em que incorremos ao não cumprir o desafio de Morgan

foi recentemente apelidado de "antropofabulação" por Cameron Buckner, visto que o erro envolve olhar para os animais a partir de uma perspectiva antropocêntrica, além de confabular os mecanismos que os humanos usam (BUCKNER, 2013).

O escândalo Clever Hans, de 1904, demonstrou o uso do Cânone de Morgan. Hans era um famoso cavalo de trote russo que encantava multidões ao mostrar a habilidade de calcular problemas matemáticos, bem como de ler alemão e notação musical, simplesmente batendo com o casco (CANDLAND, 1993; PFUNGST, 1965). Depois de muita investigação, o psicólogo experimental Otto Pfungst descobriu que Hans não estava contando ou lendo a linguagem, mas sim reagindo aos movimentos corporais de seu proprietário, von Osten. Von Osten estava inconscientemente aconselhando Hans a começar e parar de bater com o pé no momento correto, e Hans apenas aprendeu a associar os movimentos de von Osten ao comportamento correto. Hoje, o legado de Clever Hans pode ser visto nos métodos de controle usados durante o teste experimental da habilidade de um animal. Por exemplo, os pesquisadores que sabem a resposta correta usarão uma máscara de soldador, óculos escuros ou algum outro dispositivo para evitar que o sujeito seja induzido pelo olhar ou pelas expressões faciais. Outro método é usar treinadores ingênuos durante o teste.

Por volta dessa época, outros psicólogos experimentais nos Estados Unidos e na Rússia estavam interessados em descobrir princípios de aprendizagem. Nos Estados Unidos, Edward Thorndike (1874-1949) adotou o método experimental de Morgan, mas rejeitou seu apelo à introspecção. Thorndike preocupava-se com o fato de a introspecção não ser científica, porque não é observável e não podemos testar sua confiabilidade ou validade. O comportamento, por outro lado, pode ser observado e quantificado por vários observadores, de modo que Thorndike se voltou para o comportamentalismo como um modo de estudar o comportamento animal.

Thorndike focou-se na habilidade dos animais em resolver problemas. Em um conhecido conjunto de experimentos, Thorndike colocou gatos em uma variedade de caixas de quebra-cabeça e observou as estratégias que os gatos usaram para escapar. Quando colocados pela primeira vez em uma nova caixa, os gatos demoraram muito para encontrar a solução, mas depois de se tornarem mais experientes com a caixa, eles conseguiram escapar muito mais rapidamente. Thorndike descobriu que os gatos melhoraram seu tempo de reação, ignorando as ações ineficazes e realizando as úteis. Isso sugeriu a Thorndike que os gatos aprendem por tentativa e erro, e sua conclusão ajudou a reforçar a crença de que o comportamento animal pode ser totalmente explicado em termos associativos.

Embora os psicólogos tenham conseguido introduzir o tão necessário rigor no estudo das mentes dos animais, havia alguma preocupação de que eles tivessem ido longe demais, que os métodos fossem muito rigorosos e que o impulso para experimentos repetíveis e controlados não poderia ser usado para descobrir tudo o que há para saber sobre a função das mentes animais. Por exemplo, os etologistas pensavam que para entender o comportamento animal, os animais devem ser observados em seu ambiente natural. Como experimentos de laboratório esterilizados despojam os sujeitos do contexto social e ambiental, surgiu a preocupação de que alguns estudos possam ser ecologicamente inválidos.

Com a queda do comportamentalismo e a ascensão do cognitivismo na psicologia, os pesquisadores da cognição animal voltaram a investigar as mentes dos animais. Hoje, existem muitas abordagens para estudar animais experimentalmente, em laboratórios, zoológicos, salas de estar de cães, florestas, campos e oceanos. Um programa de pesquisa do *Primate Research Institute* (PRI) da Universidade de Kyoto investiga mentes de chimpanzés combinando pesquisa experimental em cativeiro com chimpanzés em Kyoto e pesquisa experimental e observacional selvagem com chimpanzés na África (MATSUZAWA *et al.*, 2006). Primeiro, o desenvolvimento físico, cognitivo e social dos chimpanzés é levado em consideração no planejamento dos experimentos, e os indivíduos são criados por suas mães, em vez de cuidadores humanos ou animais não aparentados. Além disso, o trabalho de laboratório e o trabalho de campo são sintetizados; as observações de campo são usadas para desenvolver experimentos, e os experimentos são conduzidos tanto no campo quanto no laboratório. Finalmente, o método inclui a análise das características fisiológicas e biológicas das espécies que podem estar relacionadas às habilidades cognitivas.

A pesquisa experimental no PRI usa o que eles chamam de método de “observação participante”, que se baseia na relação social triádica entre mãe, bebê e experimentador. Ao testar chimpanzés no laboratório, eles nunca são retirados de seu ambiente social natural. Em vez disso, os experimentos são trazidos para o ambiente social. À medida que um pesquisador se torna um membro desse ambiente social, ele pode realizar experimentos que são integrados às atividades diárias normais. No PRI, um pesquisador diferente está ligado a cada díade mãe-bebê, e o relacionamento deve durar por toda a vida. Acredita-se que essa relação estreita entre humanos e chimpanzés ofereça muitos benefícios. Isso torna os chimpanzés mais dispostos a se envolver nas atividades de pesquisa, para que o pesquisador possa obter um melhor entendimento do que os chimpanzés podem



e não podem fazer (em vez do que eles querem ou não fazer). Além disso, Matsuzawa afirma que o método de observação participante se sai melhor na investigação da cognição social típica da espécie do que experimentos isolados em indivíduos únicos, porque os chimpanzés do PRI não estão integrados em um ambiente social humano, mas os próprios pesquisadores se adaptam ao ambiente social do chimpanzé (MATSUZAWA, 2006a). Por fim, o vínculo entre pesquisador e sujeito permite ao ser humano interagir com seus chimpanzés “parceiros de pesquisa” numa idade ainda mais jovem, dada a relação de confiança entre pesquisador e mãe. Mãe e bebê podem aprender uma tarefa juntos, o que pode ajudar a iluminar as diferenças de desenvolvimento em habilidades específicas. Por exemplo, Inoue e Matsuzawa (2007) relatam que chimpanzés bebês são mais capazes de lembrar sequências de números em ordem do que chimpanzés adultos e humanos.

Outra forma de evitar a problemática da validade ecológica dos experimentos é realizá-los no ambiente natural do animal. Em experimentos de campo, os pesquisadores intervêm nos ambientes naturais de seus sujeitos, em vez de em um laboratório. Um tipo de experimento de campo é o experimento de reprodução, no qual os pesquisadores reproduzem uma gravação de áudio de um grito animal para testar as respostas comportamentais ao grito em vários contextos. Cheney e Seyfarth descrevem a estrutura dos experimentos de reprodução da seguinte maneira: os pesquisadores primeiro coletam gravações de vários gritos feitos por diferentes indivíduos. Em seguida, eles examinam as respostas do sujeito quando uma gravação do grito de algum indivíduo é amplificada por um alto-falante oculto (o indivíduo cuja chamada está sendo reproduzida geralmente está fora de vista em tais experimentos). O sujeito é filmado antes e depois da reprodução. Esses experimentos podem ser usados para testar como um mesmo sujeito responde a dois gritos diferentes reproduzidos no mesmo contexto, e como o sujeito responde ao mesmo grito em contextos diferentes (CHENEY; SEYFARTH, 2007). Os experimentos de reprodução foram usados em muitos contextos entre as espécies. Por exemplo, tais experimentos mostraram que os macacos-vervet respondem de forma diferente a diferentes guinchos de alerta dados para diferentes predadores (SEYFARTH *et al.*, 1980), que os babuínos grunhem para se reconciliarem após uma luta (CHENEY *et al.*, 1995), e que os babuínos se reconciliam após um parente do agressor grunhir (WITTIG *et al.*, 2007a). Os experimentos de reprodução também sugerem que os golfinhos-roaz reconhecem o assobio característico de seus parentes (SAYIGH *et al.*, 1998; JANIK *et al.*, 2006), e que os grandes chapins masculinos diferem em dimensões de personalidade (AMY *et al.*, 2010).

### 3.3 Métodos Observacionais

Enquanto no período inicial os psicólogos realizavam experimentos de comportamento animal em laboratório, etologistas como Oskar Heinroth, Konrad Lorenz, Nikolaas Tinbergen e Karl von Frisch seguiam os animais em campo para observar seu comportamento natural. O método etológico vem da biologia e leva em consideração não apenas o comportamento, mas também o contexto do comportamento, o ambiente, a fisiologia e a história evolutiva do animal. No entanto, pelo menos inicialmente, os etologistas estavam menos interessados em mecanismos cognitivos do que os experimentalistas. A etologia clássica deu origem a vários campos de estudo diferentes. A ecologia comportamental desenvolveu-se em um campo único orientado para determinar como um comportamento evoluiu e as funções desse comportamento (KREBS; DAIES, 1993; WILSON, 1975). A etologia cognitiva nasceu como um campo focado na consciência animal (GRIFFIN, 1984, 1985, 1992), embora nesta literatura a consciência seja focada tanto quanto as crenças, intenções, autoconsciência, engano e teoria da mente. O uso de Griffin da palavra "consciência" mascara seu maior interesse na cognição, dado que a maioria dos tópicos que ele discute (por exemplo, percepção, memória, cognição espacial, linguagem, uso de ferramentas, etc.) são cognitivos (GRIFFIN, 1992). Hoje, poucos pesquisadores dão boas-vindas ao rótulo de etologia cognitiva (ALLEN, 2004; SHETTLEWORTH, 2010a). Em vez disso, testemunhamos uma integração das abordagens experimentais e etológicas no estudo da cognição por cientistas em psicologia, biologia e antropologia, entre outros campos (SHETTLEWORTH, 2010a). Alguns agora usam o termo "ecologia cognitiva" para se referir ao estudo de processos cognitivos no ambiente natural de um animal (HEALY; BRAITHWAITE, 2000; REAL, 1993).

Nos primórdios da etologia, Lorenz e Tinbergen estavam interessados em analisar o conjunto complexo e rígido de movimentos que constituem um único ato. Eles postularam que tais movimentos, que chamaram de padrões fixos de ação, são inatos e causados pela existência de um mecanismo de liberação que responde a algum estímulo sensorial externo. Os etologistas estudam tais atos em vários níveis organizacionais, por exemplo nos níveis de indivíduo, díade, grupo familiar e espécie (MENZEL, 1969). Para explicar o comportamento, os etologistas seguem a sugestão de Tinbergen de que podemos distinguir entre explicações em termos

de **causas imediatas**, como mecanismo e função, e **causas últimas**, como ontogenia (os processos maturacionais envolvidos no comportamento) e evolução (TINBERGEN, 1963).

Mas antes que uma explicação para algum comportamento possa ser encontrada, o comportamento deve ser bem compreendido no contexto do comportamento normal da espécie. Assim, o etologista começará o estudo de uma espécie construindo um etograma a partir de anotações de campo feitas após muitas horas de observação (BROWN, 1975; LEHNER, 1996). Um etograma é um catálogo completo das unidades comportamentais características de uma espécie, no qual cada unidade recebe uma descrição verbal e talvez uma imagem. Surgem debates teóricos sobre a questão de como um etograma deve rotular e descrever unidades comportamentais. Os comportamentos podem ser descritos formalmente, e descrever a ação ao nível das contrações musculares, ou padrões de movimentos corporais (por exemplo, bicadas de bico no solo). Por outro lado, um etograma poderia descrever comportamentos funcionalmente e colocar o comportamento em um contexto mais amplo, referindo-se ao propósito ou consequência do comportamento, por exemplo, alimentar-se (HINDE, 1970).

Existem críticas aos métodos de descrição funcionais e formais. As descrições formais podem deixar de fora aspectos importantes do comportamento de um animal, enquanto as descrições funcionais estão sujeitas a interpretações exageradas e podem levar ao antropomorfismo, podendo confundir explicações em termos de causas finais e imediatas (*vide* ALLEN; BEKOFF, 1997). Millikan (1993) e Allen e Bekoff (1997) fornecem defesas filosóficas do emprego de descrições funcionais em etologia. Enquanto Millikan alegou que os etologistas deveriam se focar apenas no comportamento conforme descrito funcionalmente, Allen e Bekoff argumentam que a escolha entre uma descrição funcional e formal deverá variar contextualmente, dependendo de qual é mais útil. Em muitos casos, as descrições funcionais serão preferidas devido às vantagens identificadas por Hinde (1970). Por um lado, o comportamento descrito funcionalmente resultará em menos conjuntos de dados, tornando a análise de dados mais robusta. Além disso, descrições em termos funcionais são mais informativas do que as formais, uma vez que incluem informações sobre a causa do comportamento ou sobre sua consequência. Por fim, mudanças comportamentais podem ser descritas em termos de mudanças ambientais.

Independente do modo como um etologista decida descrever comportamentos, surge a questão de como individualizar um comportamento (RUSSON, *et al.*, 2007; SKINNER, 1935). As descrições de comportamento podem ser detalhistas, referindo às especificações de um comportamento; por exemplo, usar uma pedra (ou um

copo de folha, ou folhas mastigadas, ou uma mão, ou pele, etc.) para beber água de um rio. Por outro lado, os comportamentos podem ser transformados aproximadamente em unidades comportamentais maiores; por exemplo, usar uma ferramenta para beber. Se os comportamentos devem ser categorizados para refletir a maneira como a espécie organiza seu comportamento, então a identificação de comportamentos requer primeiro conhecer o esquema organizacional interno da espécie (BYRNE 1999; RUSSON *et al.*, 2007).

Enquanto os etogramas são usados para registrar comportamentos típicos, muitos pesquisadores também coletam comportamentos raros e não padronizados, bem como comportamentos que simplesmente não foram observados antes. Esses relatórios são frequentemente chamados de "incidentes" ou "relatórios qualitativos" em vez de "anedotas", a fim de evitar a conotação negativa associada ao método anedótico dos contemporâneos de Darwin. Os pesquisadores publicam incidentes quando indicam que uma espécie se envolve em comportamentos anteriormente desconhecidos. Por exemplo, quando Jane Goodall relatou ter visto chimpanzés caçando e agressões letais entre grupos, a imagem científica dos chimpanzés teve que ser significativamente revisada (GOODALL, 1986). Um exemplo de um achado observacional mais recente sugere que os chimpanzés jovens podem brincar com paus da mesma forma que crianças humanas brincam com bonecas (KAHLENBERG; WRANGHAM, 2010).

O psicólogo Richard Byrne defende o uso científico de eventos raros como uma ferramenta de pesquisa útil, escrevendo que "o registro cuidadoso e imparcial de eventos inesperados ou raros, seguido de comparação e uma tentativa de análise sistemática, não pode ser prejudicial. No pior dos casos, o exercício será substituído e tornado redundante por métodos que proporcionam maior controle; no melhor dos casos, os dados coletados podem se tornar importantes para a teoria" (BYRNE, 1997, p. 135). Ao compartilhar seus relatórios de incidentes, os pesquisadores se envolvem em projetos de pesquisa colaborativa para estudar comportamentos raros ou incomuns, como engano (WHITEN; BYRNE, 1988) e inovação (VAN SCHAIK *et al.*, 2006), e o estudo sistemático geralmente começa com observações de comportamento espontâneo (BATES; BYRNE, 2006; RUSSON; ANDREWS, 2010, 2011; VAN SCHAIK *et al.*, 2006).

#### 4. Programas de Pesquisa

Os vários programas de pesquisa em cognição animal podem ser vistos como uma tentativa de descobrir os mecanismos subjacentes envolvidos no comportamento. A abordagem funcional visa categorizar o comportamento funcionalmente (de acordo com sua função final ou imediata) e descobrir o(s) mecanismo(s) cognitivo(s) usado(s) em tais comportamentos. O problema inicial, é claro, é saber se o comportamento conta como tendo uma função específica; isso é particularmente problemático quando uma determinada função tem um mecanismo intimamente identificado com ela. Por exemplo, embora a comunicação entendida por um biólogo em termos de função final não seja nada mais do que um processo adaptativo de troca de informações entre dois indivíduos que não requer cognição, o uso do termo para descrever um comportamento pode ter implicações griceanas<sup>44</sup> para alguém, o que pode levar a inferências injustificadas sobre o mecanismo.

O comportamento animal tem sido explicado há muito tempo em termos do que às vezes se pensa serem métodos de aprendizagem associativa “simples”, como os condicionamentos clássico e instrumental, ou em termos de respostas específicas da espécie. As respostas específicas da espécie são o que às vezes é chamado de comportamento inato, mas esse termo é problemático por uma série de razões (BATESON e MAMELI, 2007; MAMELI; BATESON, 2011) e foi amplamente abandonado pelos cientistas; elas se referem, em vez disso, às predisposições demonstradas por membros típicos da espécie ou por um subconjunto da espécie (por exemplo, o guincho longo do orangotango macho). A aprendizagem associativa ocorre pela experiência de conexões entre eventos, e resulta na expectativa de que conexão continuará. No entanto, quanto mais aprendemos sobre os processos envolvidos na construção de associações, e quanto mais parece que podemos fazer com tais processos, menos “simples” e mais cognitivos eles parecem ser (CARRUTHERS, 2009, DE WIT; DICKINSON, 2009; DICKINSON, 2009; RESCORLA, 1988; SHETTLEWORTH, 2010A). Outros processos que se acredita estarem envolvidos em algumas instâncias do comportamento animal são a inferência transitiva, o aprendizado causal e a integração de caminhos. Surgem questões sobre a natureza e os requisitos cognitivos desses processos, tal como sobre sua relação com a aprendizagem associativa.

---

<sup>44</sup> N.T.: Referência ao filósofo H. P. Grice.

Os programas de pesquisa em cognição animal são numerosos demais para serem cobertos completamente aqui; felizmente, existem bons textos introdutórios de psicologia (SHETTLEWORTH, 2010a; WYNNE; UDELL, 2013) e de biologia (DUGATKIN, 2013). O que se segue é uma breve introdução a algumas áreas de pesquisa que têm sido de interesse a filósofos.

## **4.1 Comunicação**

A comunicação animal é frequentemente descrita como troca de informações entre um emissor que sinaliza para um receptor. As abordagens biológicas consideram que a comunicação ocorre quando o comportamento de um animal serve como estímulo causando uma mudança no comportamento de outro animal, independentemente de ser o comportamento do sinalizador intencional ou voluntário. Por exemplo, o inchaço genital de uma fêmea de chimpanzé, observado por um macho que inicia um cortejo, é um exemplo de comunicação. Em contextos de biologia, isso é frequentemente referido como “comunicação funcional”.

Para uma discussão da teoria e pesquisa contemporânea em comunicação animal, consulte a coleção *Animal Communication Theory: Information and Influence*, de Ulrich E. Stegmann (2013).

### **4.1.1 Sinais Expressivos e Referenciais**

Aqueles que argumentam que os sistemas de comunicação animal e a linguagem humana são homólogos (funcionalmente semelhantes devido à origem evolutiva comum) ou análogos (funcionalmente semelhantes mas com origens evolutivas diferentes) tentaram demonstrar que alguns sinais animais são referenciais em algum sentido. Um teste para comunicação referencial é ver se o comportamento é flexível, determinando se existe uma conexão meramente probabilística entre o estímulo e a resposta. Por exemplo, se houver diferentes respostas a diferentes enunciadores (por exemplo, bebê x adulto, dominante x submisso), acredita-se que isso demonstre uma flexibilidade no comportamento que sugere compreensão referencial (EVANS, 2002; TOMASELLO; ZUBERBÜHLER, 2002). Além disso, acredita-se que os chamados referenciais codificarão informações específicas sobre o predador e que os animais que ouvirem o chamado de alerta perceberão essa

informação codificada (EVANS *et al.*, 1993a; EVANS, 1997).

Marler *et al.* (1992) oferecem dois critérios que devem ser atendidos para que um sinal seja funcionalmente referencial. O critério de produção requer que todos os estímulos que eliciam o sinal pertençam a uma categoria, seja uma categoria geral, como "predadores aéreos", ou uma mais específica, como "águia". O critério de percepção afirma que a emissão do sinal referencial é por si só suficiente para provocar o mesmo comportamento que seria eliciado ao perceber o referente (MARLER *et al.*, 1992). Dados esses critérios, Marler e Evans examinaram o comportamento antipredador de galinhas garnisé e descobriram que as galinhas dão, de forma confiável, diferentes cacarejos de alerta para predadores aéreos e predadores terrestres (EVANS *et al.*, 1993b; EVANS; MARLER, 1995). Como elas também se comportam de maneira diferente com os dois predadores diferentes, Marler e Evans sugerem que os gritos de alerta se referem funcionalmente ao tipo de predador que se aproxima. Quando uma galinha emite um cacarejo ao ver um falcão, eles afirmam que a galinha está se referindo ao falcão, e não simplesmente expressando medo do falcão, ou ordenando que outras galinhas para que se abriguem, se agachem e olhem para o céu.

Gritos de alerta e outras vocalizações comunicativas que atendem a esses requisitos são encontrados em muitas espécies. Os cães da pradaria de Gunnison, por exemplo, dão diferentes guinchos de alerta para humanos, falcões e cães/coiotes. Em resposta ao guincho de alerta sobre um falcão, apenas os cães da pradaria que estão na trajetória de voo do falcão respondem, correndo para uma toca. O guincho de alerta sobre um humano desencadeia uma fuga de toda a comunidade para as tocas, enquanto o guincho de alerta sobre um cão/coiote leva todos os indivíduos a correr para a borda da toca e ficar em pé (KIRIAZIS; SLOBODCHIKOFF, 2006). Os macacos-vervet também dão guinchos de alerta para diferentes predadores. Seguindo as observações de campo do zoólogo Thomas Struhsaker (1967), Cheney e Seyfarth usaram reproduções de guinchos de alerta pré-gravadas para demonstrar que quando um alerta de leopardo soa, os macacos-vervet correm para as árvores, onde estão protegidos dos leopardos devido a sua agilidade para saltar de árvore em árvore. Quando um alerta de águia soa, os macacos olham para cima e correm para os arbustos. Quando o alerta de cobra soa, os macacos se levantam bípedes e espiam a grama ao redor (CHENEY; SEYFARTH, 1990). Outras espécies encontradas com diferentes guinchos de alerta e com comportamentos diferentes para diferentes espécies de predadores incluem macacos-diana (ZUBERBÜHLER, 2000), macacos-de-campbell (ZUBERBÜHLER,

2001) e suricatos (MANSER, 2001; MANSER *et al.*, 2001). Além disso, esquilos terrestres (OWINGS; HENNESSY, 1984), esquilos arbóreos (GREEN; MEAGHER, 1998) e mangustos-anões (BEYNON; RASA, 1989) são conhecidos por terem guinchos de alerta que distinguem entre predadores terrestres e aéreos. Acredita-se que muitos outros gritos e gestos envolvam comunicação referencial desse tipo. Por exemplo, pensa-se que os guinchos alimentares dos chimpanzés indicam não apenas a presença de comida, mas também a localização ou qualidade da comida (por exemplo, SLOCOMBE; ZUBERBÜHLER, 2005, 2006). Descobertas semelhantes foram relatadas para galinhas (EVANS; EVANS, 2007). Diz-se que os golfinhos-nariz-de-garrafa referem-se a si próprios através do seu assobio característico (JANIK *et al.*, 2006).

Em *The Expression of Emotions in Man and Animals*, Darwin argumenta que os sinais dos animais são expressões de emoções com a função de retransmitir os estados emocionais ou motivacionais do sinalizador. Esses sinais são amplamente específicos da espécie e não são usados de maneira flexível. Por exemplo, Darwin descreve o gato arqueando e sibilando para um cão como uma expressão do terror e da raiva do gato (DARWIN, 1886, p. 56). Os etólogos seguiram o exemplo de Darwin ao descrever os sinais de animais como expressões de estado emocional ou motivacional, e muitos etólogos pensaram que os sinais expressivos eram incompatíveis com os referenciais. A suposição de que os sinais dos animais são expressões de emoções levou os cientistas a se concentrarem em questões como a de se os animais comunicam o grau de excitação ou de motivação (expressão emocional analógica, e não apenas binária), e muitos estudos sugerem que alguns o fazem. No entanto, desde a década de 1980, há um crescente corpo de evidências que abala a suposição de que existe uma dicotomia entre sinais expressivos e referenciais, e alguns argumentam que os sinais animais podem informar os receptores sobre ambos o estado motivacional e objetos ou eventos externos (MARLER *et al.*, 1992; MANSER *et al.*, 2002). Por exemplo, podemos ver evidências para essa afirmação na pesquisa sobre guinchos de alerta de suricatas. Suricatas são uma variedade de mangustos africanos conhecidos por darem diferentes guinchos de alerta para predadores mamíferos, aviários e reptilianos, e eles respondem de forma diferente aos guinchos, dependendo do grau de urgência que o guincho demonstra (MANSER *et al.*, 2001). Em experimentos de reprodução, os suricatas respondem de maneira qualitativa diferente aos três diferentes guinchos de alerta e respondem de forma quantitativa diferente a diferentes níveis de urgência dentro de cada classe de guincho de alerta. Por exemplo, em resposta a uma



gravação de alerta de cobra de baixa urgência, os suricatas levantam suas caudas, se aproximam do alto-falante e farejam a área ao redor, mas eles retomarão rapidamente sua atividade anterior. No entanto, se um guincho de alarme de alta urgência for reproduzido, os suricatas continuarão o comportamento de resposta ao alarme por um tempo significativamente mais longo.

Embora seja claro que a linguagem humana pode se referir simultaneamente a eventos externos e expressar os sentimentos de um indivíduo sobre esses eventos (por exemplo, “Fogo!” vs. “Fogo?”), permanecem questões sobre a natureza da referência na comunicação animal, e se ela é a mesma natureza pela qual expressões linguísticas fazem referência. Descrever um grito como funcionalmente referencial é permanecer agnóstico sobre o papel da cognição. Além disso, isso deixa de lado a questão de saber se o ato é voluntário, intencional ou envolve representações mentais. Questões filosóficas podem ser levantadas sobre a natureza referencial dos sinais animais em termos de seu significado, atribuição de verdade, e os mecanismos envolvidos. Um tratamento filosófico da natureza dos guinchos de alerta animal defende a visão de que eles são melhor entendidos como confissões [*avowals*] neo-expressivas, autorrelatos do estado mental atual de alguém, possuindo tanto um componente de ação (a expressão) como um componente semântico (a representação) (MCANINCH *et al.*, 2009). Allen e Saidel sugerem que um trabalho empírico pode ser feito para determinar os tipos de comunicação referencial em que diferentes espécies podem se engajar, e que tal pesquisa pode nos ajudar a compreender melhor os mecanismos que operam na linguagem humana (ALLEN; SAIDEL, 1998). Por exemplo, pode-se investigar a alegação de que os assobios de assinatura de golfinhos funcionam como nomes próprios em relação a várias teorias de referência.

#### 4.1.2 Sinais Intencionais

A maioria das discussões sobre sinais intencionais na comunicação animal entra em contato com o tratamento de Daniel Dennett em algum ponto. Em sua discussão sobre os possíveis significados dos gritos de alerta dos macacos-vervet, Dennett fornece uma análise dos níveis de intencionalidade que podem estar operando (DENNETT, 1987). Dennett sugere que os pesquisadores da cognição animal se beneficiariam com a adoção da linguagem da psicologia do senso comum. Além disso, ele afirma que animais como os macacos-vervet devem ser vistos como

sistemas intencionais cujos comportamentos são previsíveis através da atribuição de crenças, desejos e racionalidade a eles; ou seja, Dennett sugere que os cientistas devem falar sobre as crenças animais (entendidas em termos de posturas intencionais, é claro). A questão, então, é que tipo de crenças os animais têm quando vocalizam: os vervets pensam sobre os efeitos de seus guinchos no comportamento dos outros, ou eles pensam sobre os efeitos de seus guinchos nos estados epistêmicos ou mentes dos outros?

A visão de Dennett é inspirada por sua interpretação da teoria do significado de H. P. Grice. Para Grice, um falante significa algo por um enunciado  $x$  se, e somente se, o falante enuncia  $x$  com a intenção de: (1) produzir uma resposta no público-alvo; (2) que o público reconheça a intenção primária do falante; e (3) que o reconhecimento pela audiência da intenção primária do falante sirva como uma razão para a audiência responder da forma como o faz (GRICE, 1957). Dada a exigência de Grice, Dennett sugere que apenas criaturas que podem manter uma crença de terceira ordem (por exemplo, eu penso que ela pensa que eu penso) serão comunicadores (DENNETT, 1987). Isso implicaria que apenas criaturas que leem mentes, isto é, aquelas que podem pensar sobre os estados mentais dos outros, são capazes de se comunicar. Isso deixa de fora alguns adultos humanos cognitivamente diversos, bebês humanos e provavelmente a maioria dos animais.

Embora critérios fortes para comunicação intencional exijam intencionalidade de terceira ordem, critérios inferiores podem ser definidos (GÓMEZ, 2007; MOORE, 2014, 2015; SPERBER; WILSON, 1986). Talvez o requisito mais básico para que um sinal comunicativo seja intencional seja estar sob o controle voluntário do sinalizador. Além da evidência neurobiológica de que alguns sinais de animais são voluntários (ALLEN; SAIDEL, 1998; PLATT; GHAZANFAR, 2010), há evidências comportamentais de que os sinais de animais podem ser usados com flexibilidade de maneiras que sugerem voluntariedade. Efeitos do público são um exemplo da flexibilidade que caracteriza alguns sinais. Galos, por exemplo, darão cacarejos de alerta com mais frequência na presença de uma galinha do que quando estão sozinhos (EVANS; MARLER, 1995), e também na presença de outro membro da espécie em vez da presença de uma codorna (KARAKASHIAN *et al.*, 1988). Outro exemplo de aparente flexibilidade no controle de vocalizações vem de observações de patrulhas de chimpanzés na fronteira, uma prática extremamente arriscada que pode resultar em morte após um encontro entre grupos. Conforme os chimpanzés se aproximam da fronteira de uma comunidade vizinha, eles ficam estranhamente silenciosos (MITANI *et al.*, 2002).

Também há evidências comportamentais de que os sinais de animais podem ser sensíveis ao contexto da mesma forma que as expressões linguísticas. Assim como alguns argumentam que o contexto pragmático da linguagem faz um trabalho semântico significativo para os humanos (STALNAKER 1999; SPERBER; WILSON, 1986), tanto a produção quanto as respostas dos sinais animais podem ser dependentes de fatores contextuais. Por exemplo, as danças de abanar das abelhas comunicam aos companheiros de ninho a localização de uma fonte de alimento, mas a dança não determina sua resposta; e se um indivíduo se lembra da localização de uma fonte preferencial de alimento, a abelha pode visitar essa fonte em vez da comunicada (GRÜTER *et al.*, 2008; GRÜTER; FARINA, 2009). E, do lado da produção, há evidências experimentais de que os orangotangos repetem um gesto quando são apenas parcialmente compreendidos, e usam gestos diferentes quando uma mensagem é mal compreendida (CARTMILL; BYRNE, 2007).

Michael Tomasello argumenta que alguns sinais dos macacos são exemplos de comunicação intencional e afirma que os sinais intencionais “são escolhidos e produzidos por organismos individuais de forma flexível e estratégica para objetivos sociais específicos, ajustados de várias maneiras para circunstâncias particulares. Esses sinais são intencionais no sentido de que o indivíduo controla seu uso de forma flexível com o objetivo de influenciar os outros” (TOMASELLO, 2008, p. 14). Para fazer um sinal intencional, Tomasello acha que é preciso ser capaz de usar o sinal de maneira flexível, estar ciente do estado de atenção do parceiro comunicativo, e ser capaz de aprender o sinal.

Embora muito do interesse na comunicação animal tenha se concentrado nas vocalizações, alguns acreditam que a comunicação gestual pode ser um lugar melhor para procurar comunicação intencional, especialmente nos grandes símios e especialmente se as teorias gestuais da evolução da linguagem estiverem corretas (CORBALLIS, 2002; ARBIB, 2002; ARBIB, *et al.*, 2008; POLLICK; DE WAAL, 2007). A comunicação gestual foi estudada em gorilas (GENTY *et al.*, 2009), orangotangos (CARTMILL; BYRNE, 2007; RUSSON; ANDREWS, 2010) e chimpanzés (TOMASELLO, 1994; PIKA; MITANI, 2006; PIKA *et al.*, 2005). Relatórios recentes afirmam que os orangotangos também usam sequências de gestos para comunicar pedidos e outras mensagens, representando ou simulando o que pretendem comunicar. Por exemplo, no contexto de um jogo de limpeza de cabeças, um jovem orangotango semilivre foi observado pegando uma folha da mão de um humano quando este se recusou a usá-la para limpar sua cabeça. O orangotango então esfregou brevemente a cabeça com a folha e, em seguida, devolveu a folha ao humano, que começou a

limpar a cabeça do orangotango com ela (RUSSON; ANDREWS, 2010). Relatórios assistemáticos de comportamento de pantomima também podem ser encontrados sobre outros grandes símios, embora programas formais de pesquisa sobre pantomima em grandes símios precisem ser mais desenvolvidos (RUSSON; ANDREWS, 2011).

### 4.1.3 Comunicação Simbólica

No século 20, houve grande interesse em ensinar sistemas de comunicação simbólica para outras espécies. As primeiras incursões nessa área foram com chimpanzés, e focaram no ensino da linguagem falada para chimpanzés criados como crianças humanas (KELLOGG; KELLOGG, 1933; HAYES; HAYES, 1951). Com a constatação de que os chimpanzés não têm o aparato vocal necessário para pronunciar palavras humanas, a pesquisa mudou para o ensino da Língua de Sinais Americana (*American Sign Language*) [LSA, doravante] e de sistemas artificiais de comunicação simbólica. O primeiro estudo, o Projeto Washoe de Beatrix e Allen Gardner, foi inicialmente relatado como um sucesso. Usando métodos explícitos de treinamento, incluindo formação, moldagem e modelagem, os pesquisadores foram capazes de treinar o bebê Washoe para formar pelo menos 132 sinais da LSA. O foco estava na produção de gestos, em vez da compreensão, e a intenção declarada dos Gardners era treinar Washoe (e, mais tarde, outros chimpanzés) em um ambiente social, imitando o ambiente de aprendizagem da linguagem das crianças tanto quanto possível. Os Gardners afirmam que “[Washoe] aprendeu uma linguagem humana natural e suas primeiras elocuições eram muito semelhantes, talvez indistinguíveis das primeiras elocuições de crianças humanas. Agora, a questão categórica, pode um ser não humano usar uma linguagem humana, deve ser substituída por questões quantitativas: quanta linguagem humana, e quão longe (ou quão perto) eles podem ir?” (GARDNER; GARDNER, 1978, p. 73).

Embora as afirmações dos Gardners sobre a linguagem dos símios estivessem sendo ecoadas por outros que trabalhavam com a linguagem dos macacos (PREMACK, 1971; PATTERSON, 1978), nem todos concordaram. O psicólogo Herbert Terrace, que usou os métodos dos Gardners para ensinar sinais LSA a um bebê chimpanzé chamado Nim Chimpsky, argumentou que os chimpanzés não estavam usando os sinais para se comunicar. Terrace concluiu que alguns dos resultados alcançados pelos Gardners poderiam ser explicados mais pelo aprendizado

associativo do que pela compreensão da semântica dos símbolos. Em seu estudo, ele tentou controlar o aprendizado associativo, e seu foco na sintaxe o fez prestar atenção à ordem dos símbolos em sequências de vários símbolos. Embora os primeiros resultados deste estudo parecessem promissores, depois de assistir a vídeos do uso do símbolo de Nim, ele percebeu que o que havia sido inicialmente visto como elocuições espontâneas muitas vezes eram imitações de elocuições recém feitas por seus treinadores. Terrace revisou filmes das elocuições de Washoe e encontrou padrões semelhantes: o professor inicia a sinalização, e o chimpanzé imita os sinais do professor. Ele também observou que o ritmo de dar e receber encontrado na comunicação criança-adulto não era reencontrado nas conversas do chimpanzé-treinador, e considerou essa diferença na pragmática como mais uma evidência de que os chimpanzés não estavam usando a linguagem (TERRACE *et al.*, 1979).

Embora os Gardners tenham defendido seus estudos contra as críticas de Terrace (GARDNER; GARDNER, 1989), outros pesquisadores tentaram controlar as interpretações alternativas de seus resultados. Premack, por exemplo, baseou-se em testes de transferência como evidência de que o chimpanzé Sara entende os símbolos que lhe ensinaram (PREMACK, 1971). Em um teste de transferência, um novo símbolo é ensinado apenas no contexto de um subconjunto do vocabulário do sujeito. Uma vez que o sujeito atinge o critério na série de ensino, um teste formal é conduzido usando novas sequências de símbolos.

A pesquisa pós-Terrace sobre comunicação simbólica se expandiu para incluir diferentes espécies, como outros macacos, golfinhos, papagaios e leões marinhos. Além disso, o foco de alguns estudos mudou da sintaxe para a semântica e da produção para a compreensão. Mais recentemente, a investigação se outras espécies podem aprender um sistema de símbolos a fim de se comunicar com os humanos foi em grande parte perdida, com o foco voltando a ser investigar as habilidades sintáticas independentemente do significado comunicativo, e as capacidades de comunicação natural dos animais.

## PESQUISA DE COMUNICAÇÃO SIMBÓLICA

Espécie	Estudo	Descrição
Chimpanzé	Kellogg e Kellogg (1933)	Co-criação de uma fêmea de chimpanzé de 7 meses e meio, Gua, com seu filho de 10 meses, Donald, por nove meses. Ambos foram treinados explicitamente no inglês falado. Embora Gua não tenha conseguido produzir linguagem, diz-se que ela compreendia 95 termos no final do estudo.
Chimpanzé	Hayes e Hayes (1951; 1952)	Uma chimpanzé fêmea Vicky foi criada desde a infância como uma criança humana por quase 8 anos. Apesar do treinamento extensivo, Vicky foi capaz de pronunciar apenas quatro palavras.
Chimpanzé	Gardner e Gardner (1971)	Ensino explícito de sinais da LSA para uma chimpanzé Washoe em um ambiente social. Washoe tinha 11 meses quando o projeto começou, e após 51 meses de treinamento ela atingiu o critério em 132 sinais.
Chimpanzé	Premack (1971)	Ensino explícito do uso de símbolos para uma chimpanzé Sarah de 6 anos em um ambiente de laboratório. Sarah foi ensinada a associar objetos, ações, classes, conectivos lógicos, etc. com fichas de plástico, e foi ensinada a produzir cadeias de símbolos que obedecem a regras sintáticas.

Chimpanzé	Rumbaugh (1973)	Ensino explícito de um sistema de lexigramas para a chimpanzé fêmea Lana de 2 anos e meio em um ambiente de laboratório mediado por computador. Lana produziu sequências de lexigramas que obedecem a regras sintáticas. Mais tarde, a performance de Lana foi considerada uma emulação do uso de símbolos humanos, porque ela falhou em compreender o aspecto referencial dos lexigramas.
Chimpanzé	Savage-Rumbaugh (1980)	Ensino explícito de um sistema de lexigramas para dois chimpanzés machos, Sherman (5 anos) e Austin (4 anos). A ênfase estava na semântica em vez da sintaxe. Sherman e Austin costumavam usar os lexigramas um com o outro para solicitar objetos.
Gorila	Patterson (1978)	Ensino explícito de sinais da LSA para uma gorila fêmea Koko em um ambiente social.
Chimpanzé	Terrace et al. (1979)	Ensino explícito de sinais da LSA para um chimpanzé macho de 2 semanas de idade, Nim Chimpsky, usando os métodos de Gardner e Gardner; falharam em replicar seus resultados.
Orangotango	Miles (1983)	Ensino explícito de LSA para um orangotango Chantek macho aculturado em um ambiente social. Chantek tinha 9 meses quando o projeto começou e continua quase 20 anos depois.

Leão Marinho	Schusterman et al. (1984)	Ensino explícito de compreensão de um sistema de comunicação gestual artificial para uma leoa marinha, Rocky, desde 1978, modelado a partir do sistema de comunicação do golfinho nariz de garrafa desenvolvido por Lou Herman.
Chimpanzé	Matsuzawa (1985)	Ensino explícito do uso de numerais para uma chimpanzé fêmea Ai em um ambiente social. Ai tinha 1 ano de idade quando chegou a Kyoto em 1977. A pesquisa continua sobre a linguagem, numérica e outras habilidades cognitivas dos chimpanzés, incluindo estudos de desenvolvimento do filho de Ai, Ayumu, usando o método de observação participante.
Bonobo	Savage-Rumbaugh (1986)	Aquisição espontânea do uso de símbolos de lexigrama em um bonobo Kanzi de 2 anos e meio, após quase dois anos observando uma tentativa explícita de ensinar sua mãe adotiva.
Golfinho Nariz de Garrafa	Herman et al. (1986)	Ensino explícito de compreensão de um sistema de comunicação gestual artificial com alguma estrutura lógica para quatro golfinhos em cativeiro, Phoenix, Akeakamai, Hiapo e Elele.
Chimpanzé	Fouts et al. (1989)	Aprendizagem social de LSA de um chimpanzé treinado Washoe para um jovem chimpanzé ingênuo Loulis.



Chimpanzé	Boysen e Berntson (1989)	Ensino explícito de numerais para uma chimpanzé fêmea aculturada, Sheba.
Papagaio Africano Cinza	Pepperberg (1999)	Modelagem social da linguagem falada usada para ensinar o papagaio Alex a vocalizar palavras em inglês. Alex foi capaz de rotular objetos por nome, cor, forma e matéria.
Orangotango	Shumaker (1997)	Ensino explícito de um sistema de comunicação de lexigramas simbólicos com alguma estrutura lógica para o orangotango Azy, em curso desde 1995.
Border Collie	Pilley e Reid (2011)	Ensino explícito dos nomes próprios únicos (falados em inglês) de 1022 objetos para Chaser, cujo treinamento começou por seus cuidadores quando ela tinha 8 semanas de idade.

Os defensores desse programa de pesquisa argumentam que os estudos revelam algo sobre a relação entre a linguagem e a mente, a evolução da linguagem humana, e os papéis desempenhados pelo desenvolvimento e estruturamento da linguagem humana (LLOYD, 2004). No entanto, para os críticos, esses estudos simplesmente fornecem mais evidências a favor do poder de associação e da capacidade dos humanos de treinar animais para fazer quase qualquer coisa. Há uma vasta literatura sobre esses estudos, com críticos (PINKER; BLOOM, 1990; PINKER, 1994; CHOMSKY, 1980), bem como defensores (LLOYD, 2004; GREENFIELD, 1991; SAVAGE RUMBAUGH et al., 1998).

Uma área de controvérsia tem a ver com se os animais que usam com sucesso algum aspecto da linguagem humana o estão usando como linguagem ou, em vez disso, estão envolvidos em comunicação simbólica. Foram oferecidas pelo menos três demarcações diferentes entre a linguagem e outros sistemas de comunicação simbólica. De acordo com o programa linguístico original de Noam

Chomsky, usar a linguagem é incorporar certos princípios estruturais, e todos os usuários da linguagem são capazes de produzir um número potencialmente infinito de sequências gramaticais por meio de incorporação recursiva (CHOMSKY, 1968). O antropólogo linguístico Charles Hockett identificou até dezessete características de design que ocorrem em todas as línguas humanas, incluindo semanticidade, discricção e arbitrariedade (HOCKETT, 1977). Mais recentemente, Hauser, Fitch e Chomsky (2002) argumentam que o mecanismo que permite o pensamento recursivo é o requisito cognitivo central para a linguagem e é uma característica dos sistemas de comunicação humana não encontrados em outras espécies. No entanto, pesquisas com estorninhos europeus descobriram que podemos treinar pássaros para discriminar uma gramática recursiva entre sequências de sons de estorninhos (GENTNER *et al.*, 2006), e que há semelhanças entre a linguagem humana e o canto dos pássaros nas dimensões cognitiva, neurológica, genômica e comportamentais (BOLHUIS *et al.*, 2010).

Chomsky era um crítico vocal dos primeiros estudos da linguagem animal, especialmente das afirmações feitas por alguns pesquisadores de que os macacos haviam adquirido a linguagem. Para Chomsky, a linguagem requer sintaxe, que ele afirmava estar faltando em todos os sistemas de comunicação dos macacos. Além disso, treinar um macaco para usar símbolos é um processo trabalhoso, enquanto as crianças aprendem a linguagem sem esforço. A linguagem é inata, de acordo com Chomsky, de modo que, se os macacos tivessem a capacidade de aprender a linguagem, eles fariam sem intervenção humana (CHOMSKY, 1968). Chomsky, frequentemente, afirma sua crítica como um argumento *a priori* contra a linguagem animal: “se um animal tivesse uma capacidade tão biologicamente sofisticada quanto a linguagem, mas de alguma forma não a tivesse usado até agora, seria um milagre evolucionário, como encontrar uma ilha de humanos que poderiam ser ensinados a voar” (LLOYD, 2004, p. 585).

Outro argumento que Chomsky ofereceu contra a linguagem animal é baseado na dissimilaridade entre os sistemas de comunicação animal e a linguagem humana. Ele escreve: “A questão de saber se outros sistemas são 'como' a linguagem humana é uma questão sobre a utilidade de uma certa metáfora” (CHOMSKY, 1980, p. 434); e ele argumenta que os princípios estruturais, a maneira de uso, e o desenvolvimento ontogenético do uso de símbolos por macacos, são algo tão diferente da linguagem humana que qualquer analogia entre os dois seria muito fraca. Aqueles que defendem o sistema de comunicação simbólica animal como linguagem questionam Chomsky neste ponto e enfatizam as semelhanças entre os dois sistemas de comunicação.

Dadas as descobertas da genética, a capacidade biológica para a linguagem pode ser descrita com mais precisão como uma coleção de capacidades biológicas, algumas das quais compartilhamos com outras espécies. O gene FOXP2, descobriu-se, desempenha um papel na produção da fala, e alguns afirmam que foi fundamental para o desenvolvimento da linguagem em humanos. O gene FOXP2 também é expresso na mesma parte do cérebro em tentilhões-zebra, e foi relatado que filhotes de tentilhões com FOXP2 reduzido são prejudicados em sua capacidade de aprender a cantar (HAESLER *et al.*, 2007).

#### 4.1.4 Comunicação Gestual

Enquanto a maioria das discussões sobre comunicação animal enfoca a comunicação vocal, a comunicação animal também pode ocorrer através de outras modalidades. A dança da abelha é um exemplo de método postural de comunicação. Os grandes símios se comunicam por meio de gestos, como pantomima (RUSSON; ANDREWS, 2011). A ideia de que a linguagem humana evoluiu a partir de movimentos corporais como gestos, mímica e dança foi promovida por Michael Corballis (1992, 2002) e Merlin Donald (1991) como a teoria gestual de aquisição da linguagem. Como, para os primatas, o movimento corporal está sob controle cortical voluntário em maior extensão do que as vocalizações, pode ser que nossos ancestrais hominídeos, assim como nossos primos macacos, usem gestos e postura para se comunicar intencionalmente. O neurocientista Michal Arbib apoia a teoria gestual da evolução da linguagem, por meio do recurso ao sistema de espelho, que é um sistema neural encontrado em humanos e outros primatas ativo tanto quando testemunha-se outro engajando-se em uma ação, quanto também quando alguém próprio se engaja nessa ação (ARBIB, 2005).

Tomasello afirma que os chimpanzés se comunicam intencionalmente apenas por meio de gestos, porque ao gesticular, mas não ao vocalizar, os macacos monitoram o olhar de parceiros comunicativos (LEAVENS; HOPKINS, 1998) e consertam as tentativas de comunicação fracassadas repetindo uma mensagem ou elaborando sobre ela (LIEBAL *et al.*, 2004, LEAVENS *et al.*, 2005).

Observou-se que os corvos usam comunicação gestual, em seus movimentos de cabeça e bico que indicam a presença de objetos como musgo ou galhos para seu parceiro (PIKA; BUGNYAR, 2011). Também foi observado que elefantes usam a comunicação postural, como quando orientam o corpo para indicar para onde

querem ir em seguida (POOLE; GRANLI, 2011). Além disso, alguns elefantes entendem o apontar de humanos (SMET; BYRNE, 2013).

## 4.2 Leitura de Mentes ou Teoria da Mente

Como os humanos, muitas espécies são animais sociais que, além de navegar em um mundo físico, também devem navegar em um mundo social. Na década de 1970, foi sugerido que, para prosperar em um mundo social competitivo, os indivíduos se beneficiariam em ter alguma compreensão da mente dos outros (HUMPHREY, 1976, 1978), com Alison Jolly sugerindo que conhecer outras mentes ajuda membros de grandes grupos sociais a cooperar (JOLLY, 1966). Humphrey escreveu: “[...] atrevo-me a sugerir que se o conhecimento de um rato sobre o comportamento de outros ratos fosse limitado a tudo o que os comportamentalistas descobriram sobre os ratos até agora, o rato mostraria tão pouca compreensão de seus companheiros que iria estragar desastrosamente todas as interações sociais em que se envolveu; as perspectivas para um homem igualmente limitado seriam ainda mais sombrias” (HUMPHREY, 1978, p. 60). A ideia aqui é que o conhecimento de outras mentes pode oferecer um valor adicional em relação ao conhecimento dos padrões de comportamento alheios.

O termo “teoria da mente” foi introduzido pelos psicólogos David Premack e Guy Woodruff em torno dessa época. A questão específica na qual Premack e Woodruff estavam interessados era se o chimpanzé atribui crenças e desejos a fim de prever e explicar o comportamento, algo que eles presumiram que os humanos fazem. Com efeito, Premack e Woodruff queriam saber se um chimpanzé é um teórico da ação humeano que entende o comportamento dos outros como sendo causado por atitudes proposicionais. Assim, embora definissem a teoria da mente como a capacidade de prever e explicar o comportamento atribuindo estados mentais, eles estavam mais focados em saber se os chimpanzés se engajam no raciocínio sobre crenças e desejos. Premack e Woodruff tentaram determinar se Sarah, o mesmo chimpanzé do projeto de comunicação simbólica de Premack, tem uma teoria da mente. Para examinar se Sarah entende o que os outros acreditam, eles usaram o seguinte paradigma: Sarah viu vídeos de humanos tentando resolver certas tarefas (por exemplo, adquirir bananas fora de alcance, aquecer uma sala fria acendendo um aquecedor) e ela deveria escolher entre uma série de fotos apontando a solução (PREMACK; WOODRUFF, 1978). Como Sarah escolheu a

fotografia correta em um nível acima do aleatório, Premack e Woodruff concluíram que ela tem uma teoria da mente. Eles alegaram que Sarah deve ter atribuído “pelo menos dois estados mentais ao agente humano, a saber, intenção ou propósito de um lado, e conhecimento ou crença do outro” (PREMACK; WOODRUFF 1978, p. 518).

Em comentário a esse estudo, foi apontado que Sarah poderia ter usado outros métodos para resolver os problemas. Ela poderia, por exemplo, estar atenta ao objetivo dos agentes, em vez de seus estados mentais (que é a interpretação que Premack agora endossa (PREMACK; PREMACK, 2003). A maioria dos comentaristas não ficaram convencidos com o design do estudo, e vários sugeriram metodologias alternativas para examinar a questão. Uma sugestão era exigir que o sujeito resolvesse um problema de coordenação. Para ter sucesso em um problema de coordenação, o sujeito teria que alterar seu próprio comportamento na expectativa do que outro faria (BENNETT, 1978; DENNETT, 1978; HARMAN, 1978). Dennett sugere que um problema de boa coordenação pode exigir que o sujeito considere a crença falsa de outro, de modo que o comportamento previsto seja incomum, como um comportamento que só seria exibido se o agente tivesse uma crença falsa. Um problema de coordenação de crença falsa evitaria, assim, interpretações alternativas relacionadas com a identificação do objetivo do agente, ou com a realização de associações com situações semelhantes do passado. O comportamento realizado por um agente que tem uma crença falsa não alcançará o objetivo do agente e provavelmente não será algo que o sujeito testemunhou anteriormente. O principal problema com essa sugestão, observa Dennett, é como determinar o conteúdo das previsões que um chimpanzé pode fazer.

Dadas as dificuldades associadas ao desenvolvimento de um bom teste não verbal para leitura mental (DENNETT, 1983), a sugestão de Dennett foi aceita por pesquisadores interessados em estudar a teoria da mente em crianças (WIMMER; PERNER, 1983). Wimmer e Perner aceitaram a definição de teoria da mente de Premack e Woodruff e se interessaram pela questão do estágio em que as crianças pequenas adquirem uma teoria da mente. Para responder a essa questão, eles projetaram o *teste da crença falsa*, que se tornaria um teste padrão para a teoria da mente. As crianças assistiram a um show em que um boneco chamado Maxi guarda um pedaço de chocolate em uma caixa antes de sair da sala. Enquanto Maxi está fora, sua mãe encontra o chocolate e o leva para um armário. Maxi retorna à cena, o show é interrompido e as crianças são convidadas a prever onde Maxi irá procurar seu chocolate. Se a criança disser que Maxi vai olhar no armário, ela é reprovada no teste e, portanto, mostra que não tem uma teoria da mente. Se a

criança diz que Maxi vai olhar na caixa, ela passa; passar na tarefa mostra que a criança tem uma teoria da mente, porque ela demonstra que pode atribuir estados mentais para prever o comportamento de Maxi.

Esse programa de pesquisa foi intimamente associado a um debate sobre psicologia de senso comum, entre a psicologia de senso comum como teoria (a visão de que o conhecimento humano de outras mentes é teórico por natureza) e a psicologia de senso comum como simulação (a visão de que nosso conhecimento de outras mentes depende do uso de nossa própria mente como modelo). Devido a esse debate, os filósofos começaram a substituir a expressão "leitura da mente" por "teoria da mente", para ser mais inclusivo. No final da década de 1990, houve uma aceitação crescente de que tanto a teoria-teoria quanto a teoria da simulação estavam parcialmente certas e parcialmente erradas, e isso culminou em uma aceitação geral de algum tipo de teoria híbrida (NICHOLS; STICH 2003; GOLDMAN, 2006). Esses argumentos fazem uso de dados empíricos tanto da literatura sobre desenvolvimento quanto da cognição animal.

Durante esse período, houve algumas tentativas de revelar a leitura mental em animais através paradigmas não-verbais, o que não obteve muito sucesso (HEYES, 1998). Dadas as discordâncias teóricas e de definição subsequentes, alguns pesquisadores concluíram que "o rótulo genérico 'teoria da mente' realmente cobre uma ampla gama de processos de cognição social" (TOMASELLO *et al.*, 2003b, 239). O programa de pesquisa em leitura mental animal posteriormente mudou de tentativas de chegar a um teste de crença falsa não verbal para questões mais específicas sobre as capacidades cognitivas, como a compreensão dos estados perceptivos (HARE *et al.* 2000), objetivos (ULLER, 2004) ou intencionalidade (TOMASELLO, 2005) dos outros.

#### **4.2.1 Leitura das Percepções da Mente [*Mindreading Perceptions*]**

Tem havido muito interesse em examinar a compreensão dos primatas sobre os estados perceptivos. A abordagem padrão na investigação começa com duas suposições: que estados perceptivos, como estados de crença, estão ocultos do ponto de vista do observador e, portanto, devem ser postulados como entidades teóricas; e que conhecer os estados perceptivos de alguém, como conhecer os estados de crença de outra pessoa, facilita predições precisas sobre o comportamento futuro. Evidências etológicas de que os chimpanzés monitoram o olhar e modificam

seu comportamento quando estão visíveis para os outros (PLOOIJ, 1978; WHITEN; BYRNE, 1988; GOODALL, 1986) foram tomadas como fornecendo evidências de que os chimpanzés podem atribuir estados perceptivos a outros. Como resultado, os pesquisadores experimentais decidiram desenvolver estudos para determinar se os chimpanzés entendem a visão.

Os resultados dos primeiros estudos de laboratório foram mistos. A pesquisa de David Premack sugeriu que os chimpanzés entendem a visão (PREMACK; PREMACK, 2003), enquanto os estudos de Povinelli e Eddy (1996) contestaram essa conclusão. Estudos posteriores sugeriram que os chimpanzés entendem tanto a visão quanto a intencionalidade (HARE *et al.*, 2000, 2001). Na configuração experimental de Hare *et al.*, um chimpanzé subordinado e um dominante são soltos em uma sala com isca de comida. Normalmente, se ambos os animais puderem ver a comida, ou se verem um ao outro testemunhando a isca, o animal subordinado evitará a comida e permitirá o acesso ao dominante. No entanto, nesses experimentos, quando a comida é ocluída da visão do dominante, o subordinado se aproxima dela. Somente se o dominante puder ver a comida ou a isca, o subordinado a evitará. Os animais estão em lados opostos da sala, então o subordinado tem que considerar a perspectiva visual do dominante para julgar corretamente se ele pode ver a comida ou não. Como parece que o subordinado é capaz de fazer julgamentos diferentes sobre se deve ou não buscar o alimento com base apenas em se ele é visível para o dominante, este estudo indica que os macacos entendem o estado mental de ver.

Povinelli e Vonk (2004) criticam o estudo de Hare *et al.*, sugerindo que a natureza ecológica do estudo (usando o comportamento de competição alimentar do repertório natural do sujeito) é uma fraqueza dele, e não uma força conforme os autores acreditavam, porque os chimpanzés poderiam ter feito inferências com base no comportamento observado no passado. Em resposta, os autores afirmam que levaram em consideração todas as explicações alternativas possíveis para o comportamento do subordinado, fazendo uma inferência para a melhor explicação para o argumento de que o subordinado entende o que o dominante vê (TOMASELLO *et al.*, 2003a, 2003b; HARE *et al.*, 2006 )

Em uma revisão da literatura, Call e Tomasello (2008) concluem que há ampla evidência de que os chimpanzés entendem os objetivos, intenções, percepções e conhecimentos dos outros, mas que não há nenhuma evidência experimental de que eles entendam a crença falsa. A evidência para esta afirmação final vem de estudos recentes, como um estudo de competição de comida com turnos que sugere que os chimpanzés podem entender o estado de conhecimento dos outros, mas

não seu estado de crença (Kaminski *et al.* 2008). Neste estudo, dois chimpanzés se revezam apontando para um entre três baldes opacos para ganhar recompensas de comida que podem estar escondidas dentro. Em uma condição, o chimpanzé sujeito observou dois baldes sendo iscados, e o chimpanzé concorrente observou apenas um balde sendo iscado. Quando o chimpanzé sujeito teve permissão para escolher primeiro, ele não mostrou preferência por nenhum dos baldes com isca. No entanto, quando o chimpanzé sujeito teve que escolher em segundo (sem ter a oportunidade de observar a escolha do chimpanzé concorrente), o sujeito escolheu com mais frequência o balde iscado que o concorrente não tinha visto. Os autores concluem que os chimpanzés às vezes sabem o que os outros sabem. Eles então usaram uma variação desse experimento para testar as crenças falsas. O sujeito viu o chimpanzé competidor sendo enganado sobre a localização real do alimento, mas não foi capaz de fazer uso dessa informação para prever a primeira escolha do competidor. Isso sugere que os chimpanzés eram incapazes de distinguir entre as crenças verdadeiras e falsas dos outros. Enquanto ao mesmo tempo os autores afirmam, “Esses resultados sugerem que, pelo menos em algumas situações, os chimpanzés sabem o que os outros sabem, no sentido de ter visto” (KAMINSKI *et al.*, 2008, p. 229).

#### 4.2.2 O Problema Lógico e a Parcimônia

Várias preocupações foram levantadas sobre o desenho dos estudos que testam a atribuição de crenças e percepções por grandes símios. Daniel Povinelli e seus colegas desafiaram os paradigmas, argumentando que em cada caso o desempenho dos sujeitos pode ser explicado por terem uma teoria do comportamento em vez de uma teoria da mente, e isso ficou conhecido como "o problema lógico" (POVINELLI; VONK, 2004). Eles afirmam que em todos os estudos que foram feitos até agora, os chimpanzés poderiam usar uma regra complementar de leitura de comportamento  $S \rightarrow C$  para prever o comportamento, em vez de uma regra de leitura mental  $S \rightarrow M \rightarrow C$  (onde  $S$  é a pista situacional,  $C$  é o comportamento previsto, e  $M$  é o estado mental). Povinelli e Vonk sugerem que tanto os humanos quanto os chimpanzés têm uma teoria do comportamento, mas os humanos também têm uma teoria da mente, e ser um leitor de mentes requer também ser um leitor de comportamento.

Uma resposta a este desafio é sugerir que, uma vez que a leitura de mente permite fazer previsões de comportamento em novas situações, podemos encontrar



evidências de leitura de mente quando vemos que um preditor faz uma previsão a partir de sua experiência pessoal, em vez de sua experiência observando os outros agir. Essa capacidade de fazer novas previsões é o valor agregado da leitura de mente em vez da mera leitura do comportamento. Assumindo essa posição, Povinelli e Vonk sugerem que o paradigma de projeção de experiência proposto por Cecilia Heyes evitaria o problema lógico, porque envolve pedir a um chimpanzé que perceba que uma situação nova (a saber, usar um balde vermelho na cabeça) estava associada a uma experiência mental particular (a saber, não ser capaz de ver). Esse teste exigiria que o sujeito fizesse uma inferência não do comportamento observado para o novo comportamento, mas da experiência introspectiva própria [*of the self*] para a experiência mental do outro. Um chimpanzé que passou neste teste seria capaz de prever que outro chimpanzé que usasse o balde vermelho não seria capaz de ver. Ou seja, o chimpanzé teria que inferir alguma variável intermediária entre o comportamento observado e a ação, e para Povinelli e Vonk isso é evidência suficiente de leitura de mente (BUCKNER, 2014). Mais recentemente, pesquisadores executaram versões do teste de óculos em chimpanzés (KARG *et al.*, 2015) e corvos (BUGNYAR *et al.*, 2016) e, em ambos os casos, descobriram que os animais são capazes de passar na tarefa.

No estudo com corvos, os sujeitos têm a experiência de ver uma sala adjacente através de um olho mágico, observando comida sendo escondida. Eles então são liberados na sala, e nesse momento apanham a comida. Depois de ter essa experiência, os pesquisadores investigaram o comportamento dos corvos de esconder alimentos na sala sob três condições: com uma janela transparente, sem uma janela transparente e sem olho mágico, e sem uma janela transparente e com olho mágico. Eles descobriram que o comportamento de esconder dos corvos era o mesmo nas condições com a janela transparente e com o olho mágico, e significativamente diferente nas condições sem a janela transparente. Os autores do estudo concluem que este estudo fornece evidências de uma teoria da mente em corvos (BUGNYAR *et al.*, 2016). No estudo com chimpanzés, os indivíduos são expostos a caixas de alimentos com tampas com diferentes propriedades transparentes, e os chimpanzés obtêm uma experiência na qual aprendem que uma tampa que parece opaca é, na verdade, transparente de outra perspectiva. Quando os chimpanzés competem por comida com um agente humano, eles preferem tirar comida de caixas opacas em vez de caixas transparentes, e também de caixas opacas em vez de caixas que apenas parecem ser opacas, mas com as quais o sujeito teve experiência anterior, e por isso sabe que a partir da visão do agente humano elas são transparentes

(KARG *et al.*, 2015). Dadas as descobertas desses dois estudos, parece que o problema lógico foi superado.

No entanto, há uma interpretação mais forte do problema lógico que pode não ser superada por esses estudos. Em uma interpretação mais forte, é necessário haver evidências de que a variável interveniente é propriamente mental, em vez de comportamental (LURZ, 2011). Levantando uma preocupação ao longo dessa linha, Kristin Andrews (2005) argumentou que, no teste de óculos proposto, o chimpanzé pode passar no teste por entender que o balde vermelho é o balde que impede a habilidade de **fazer** as coisas, em vez de ser o balde que atrapalha a habilidade de alguém para **ver** coisas; o comportamento do chimpanzé é consistente com uma regra de leitura de comportamento, bem como com uma de leitura de mente. Uma vez que a previsão que o chimpanzé participante seria solicitado a fazer seria sobre o comportamento do outro chimpanzé (como pedir comida para uma pessoa), o chimpanzé participante pode resolver essa tarefa percebendo que o balde vermelho impede as habilidades dos indivíduos de alcançar o objetivo específico. Em vez de generalizar a partir de sua própria experiência mental, o sujeito chimpanzé bem-sucedido pode generalizar a partir de sua própria experiência física. Andrews prossegue, argumentando que essa leitura mais forte do problema lógico torna a questão da leitura de mente do chimpanzé algo que não está sujeito à investigação científica, e transforma o problema lógico nos problemas céticos gerais sobre outras mentes (ANDREWS, 2015).

Em uma tentativa de evitar esses tipos de explicações alternativas, que Lurz chama de “hipóteses complementares de leitura de comportamento”, Lurz tenta inventar novos paradigmas de pesquisa para animais não humanos, de modo que passá-los não estaria sujeito a uma hipótese complementar de leitura de comportamento. Ele sugere que a leitura mental do macaco pode ser testada diretamente usando uma versão do teste de aparência-realidade. Lurz aponta que, considerando a maneira como um objeto aparece para um indivíduo, pode-se prever melhor o comportamento futuro desse indivíduo em relação a esse objeto do que se considerássemos as propriedades reais do objeto. Seguindo esse insight, ele sugere que os pesquisadores deveriam examinar se os chimpanzés levam em consideração como os objetos aparecem para os outros. Em uma revisão desse livro, Andrews (2012b) argumenta que os experimentos propostos por Lurz também falham em evitar o problema lógico, pois existem descrições comportamentais que podem ser oferecidas para um desempenho bem-sucedido nessas tarefas também.

Povinelli e colegas também sugerem outro valor agregado à leitura de mente, a saber, o de que o estado mental pode ser usado para reinterpretar o comportamento observado. A Hipótese de Reinterpretação afirma que representar os estados mentais de outras pessoas tem a função primária, não de prever um novo comportamento, mas, sim, de fornecer uma descrição causal do comportamento que pode ser previsto sem apelar para os estados mentais (POVINELLI *et al.*, 2000). Eles escrevem: “[...] a evolução dos estados intencionais de segunda ordem pode ter permitido aos humanos reinterpretar comportamentos sociais existentes e extremamente complicados que evoluíram muito antes de nós [...] uma vez que este novo dispositivo representacional entre em cena, pode haver efeitos em cascata sobre aspectos maiores do sistema – neste caso, a cultura material e social, incluindo pedagogia e ética [...]” (POVINELLI *et al.*, 2000, 533). Dada essa assunção, também se pode examinar a existência desses efeitos em cascata em outras espécies, a fim de se obter evidências de leitura da mente. No entanto, Povinelli está convencido de que os grandes símios não humanos não podem raciocinar sobre inobserváveis, e não entendem causalidade (POVINELLI; DUNPHY-LELII, 2001). Isso explica o pessimismo de Povinelli e seus colegas em relação aos chimpanzés serem leitores de mentes.

Em vez de pensar que um único experimento possa servir como evidência para a leitura de mente dos chimpanzés, outros filósofos passaram a considerar um corpo maior de evidências. Elliot Sober sugere que podemos obter evidências de uma variável interveniente mentalista em animais conduzindo um estudo de duas alas usando dois tipos diferentes de estímulos que eliciam dois tipos diferentes de comportamento, que são unificados para um sujeito mentalizador, mas não para um sujeito que lê o comportamento. Ele nos pede para considerar a execução de uma versão dos estudos de Melis *et al.* (2006) em que os chimpanzés são convidados a roubar alimentos de humanos em duas condições relacionadas. Em uma condição, os chimpanzés podem alcançar através de tubos opacos ou transparentes para obter alimento e, em outras condições, os chimpanzés podem levantar um alçapão barulhento ou silencioso. Alcançar tubos opacos e levantar portas silenciosas resulta em um roubo bem-sucedido. Sober argumenta que, de acordo com a hipótese de leitura da mente, passar por uma ala dessa tarefa deve aumentar a probabilidade de passar pela outra ala dessa tarefa, enquanto para a hipótese de leitura de comportamento deve haver uma triagem das duas condições, sem correlação entre as tarefas (SOBER, 2015). Outras objeções ao desafio do problema lógico vêm de Marta Halina (2015), que argumenta que o problema lógico é uma forma de ceticismo que não entra na investigação empírica. Fletcher e Carruthers (2013) argumentam

que a hipótese de leitura do comportamento não é falsificável.

O movimento para considerar um corpo de pesquisa como evidência de leitura de mente às vezes é expresso em termos de parcimônia (WHITEN, 1995). Por exemplo, Logan Fletcher e Peter Carruthers (2013) revisam as evidências de sucessos e fracassos em tarefas de leitura de mente de chimpanzés e, apelando para considerações de parcimônia, concluem que as evidências sugerem que os macacos leem desejos, percepções e conhecimento. A hipótese de que os chimpanzés leem a mente unifica um grande corpo de evidências. No entanto, considerações de parcimônia têm sido usadas tanto para concluir que os animais leem mentes, quanto para concluir que não o fazem. Isso levou a uma investigação que enfoca o papel da parcimônia nos debates sobre leitura de mente. Sober investiga a parcimônia evolutiva e a parcimônia de caixa preta e oferece sugestões sobre o tipo de evidência de que precisamos para avançar no debate (2015). Outras investigações sobre o papel da parcimônia na investigação da leitura de mente animal vêm de Hayley Clatterbuck (2015), que argumenta que é mais simples para os chimpanzés ler a mente do que ler o comportamento, porque um modelo de leitura da mente tem menos parâmetros ajustáveis, dado que a atribuição do estado mental pode unificar vários inputs e outputs; e de Simon Fitzpatrick (2009), que oferece uma análise do papel que a simplicidade desempenha em ambos os lados do debate.

#### **4.2.3 Além dos Chimpanzés: Leitura de Mente em Outras Espécies**

Desenvolvimentos recentes no estudo de leitura de mente em bebês humanos fornecem métodos adicionais para estudar a crença falsa em outras espécies. Apesar da alegação amplamente difundida de que as crianças não desenvolvem uma teoria da mente até cerca de quatro anos de idade (WELLMAN *et al.*, 2001), os pesquisadores que usam tarefas de resposta espontânea afirmam que a compreensão de crenças falsas se desenvolve na infância (*vide* BAILLARGEON *et al.*, 2010). Pesquisas preliminares usando os mesmos métodos usados em estudos com bebês não encontraram evidências de compreensão de crenças falsas em macacos (RUIZ, 2010; MARTIN; SANTOS, 2014). No entanto, usando um método semelhante, Claudia Uller descobriu que os chimpanzés respondem como bebês humanos a estímulos que os pesquisadores adultos interpretam como exemplo de comportamentos direcionados a um objetivo (ULLER, 2004; GERGELY *et al.*, 1995).

Vários estudos sugerem que os corvídeos demonstram habilidades cognitivas sociais semelhantes às dos macacos (BUGYNAR *et al.*, 2007; DALLY *et al.*, 2006; EMERY; CLAYTON, 2004). Por exemplo, a pesquisa sobre o comportamento dos *scrub-jay* de esconder alimentos mostra que os indivíduos que furtaram o alimento de outro *scrub-jay* no passado vão esconder novamente a comida quando outro *scrub-jay* observar o escondimento inicial, mas não o farão se o escondimento inicial não foi observado (EMERY; CLAYTON, 2004). *Scrub-jays* ingênuos não refazem o escondimento. Emery e Clayton sugerem que os *scrub-jays* que refazem o escondimento estão engajados na projeção de experiência: "eles relacionam informações sobre sua experiência anterior como furtador com a possibilidade de roubo futuro por outro indivíduo, e modificam sua estratégia de recuperação de forma adequada" (EMERY; CLAYTON, 2004, 1905).

Estudos sobre compreensão perceptual foram conduzidos em macacos-rhesus (FLOMBAUM; SANTOS, 2005). Da mesma forma que os estudos do chimpanzé e do *scrub-jay*, esses experimentos estabelecem uma situação competitiva naturalística em que o sujeito tem que prever o comportamento de um competidor. Em uma versão desse estudo, macacos-rhesus da ilha de Cayo Santiago foram colocados contra competidores humanos em uma tarefa de pilhagem; dois experimentadores se aproximavam de um macaco solitário e cada um se situava de maneira diferente, de modo que o macaco ficasse visível para um experimentador, mas não para o outro. Ambos os experimentadores tinham uma uva. Flombaum e Santos descobriram que os macacos eram mais propensos a roubar uvas do experimentador que não conseguia vê-los. Eles encontraram resultados semelhantes para audibilidade; quando tinham a opção de roubar uma uva em uma caixa transparente coberta de sinos, ou uma uva em uma caixa transparente sem ruídos, os macacos preferiam a comida silenciosa quando ninguém estava olhando para eles. No entanto, quando ficou óbvio que o macaco estava sendo observado, não houve preferência por roubar uvas silenciosas em vez de ruidosas (SANTOS *et al.*, 2006).

Embora os chimpanzés mostrem alguma sensibilidade às intenções e objetivos, os cães domésticos podem estar ainda mais sintonizados com as intenções dos humanos. Os cães são capazes de usar o olhar de um humano para determinar onde o alimento está escondido, uma habilidade não demonstrada pelo chimpanzé (HARE *et al.*, 1998; HARE; TOMASELLO, 1999; MIKLOSI; TOPAL, 2004; BRAUER *et al.*, 2006). Confira Hare e Woods (2013) para um resumo da pesquisa sobre cognição canina. Os cães parecem ser sensíveis ao olhar dos humanos e frequentemente fazem contato visual antes de começar a brincar. Uma explicação

para a acuidade social dos cães é a de que, ao selecionar características que tornam os cães melhores companheiros humanos, os humanos inadvertidamente criaram cães que são mais capazes de passar em tarefas de teoria da mente (HARE *et al.*, 2002).

Alguns pesquisadores que trabalharam de perto com os golfinhos nariz de garrafa acham que o conjunto geral de pesquisas com golfinhos sugere que eles leiam a mente. Os golfinhos cativos são capazes de passar por tarefas de escolha de objetos, nas quais um informante humano aponta para indicar onde a comida está escondida. Os golfinhos podem usar esta pista para acessar a comida (HERMAN *et al.*, 1999; TSCHUDIN *et al.*, 2001). Os golfinhos também passam por tarefas de autorreconhecimento no espelho, e Diana Reiss acha que isso oferece evidências de que os golfinhos também têm alguma capacidade de leitura de mente relacionada à empatia, porque as crianças humanas desenvolvem empatia ao mesmo tempo que se reconhecem nos espelhos (REISS, 2012).

#### 4.3 Autorreconhecimento em Espelho

Algumas pesquisas buscam explorar o que os indivíduos sabem sobre suas próprias mentes. Uma área de muita atenção tem sido o autorreconhecimento em espelho [*Mirror self-recognition*, MSR]. Nesse paradigma, desenvolvido pelo psicólogo Gordon Gallup, os sujeitos são sub-repticiamente marcados e, em seguida, recebem um espelho. “Passar” no teste MSR envolve tocar a marca com mais frequência quando há um espelho disponível do que quando não há. Gallup argumentou que passar no MSR implica que o animal tem um conceito de eu [*self*] (GALLUP, 1970), embora outros contestem essa afirmação. Embora antes fosse considerado um comportamento raro, limitado a alguns dos grandes símios, hoje muitas espécies foram estudadas e pelo menos alguns resultados positivos foram relatados para as seguintes espécies:

<b>Espécie</b>	<b>Estudo</b>
Chimpanzés	Lin et al., 1992. Swartz e Evans, 1991.
Gorilas	Shumaker e Swartz, 2002.
Orangotangos	Swartz et al., 1999

Golfinhos nariz de garrafa	Marino et al., 1994; Reiss e Marino, 2001.
Elefantes asiáticos	Plotnik et al., 2006.
Pegas	Prior et al., 2008.
Macacos-rhesus	Rajala et al., 2010.

Muitas outras espécies falharam em mostrar comportamento direcionado pelo espelho, incluindo algumas espécies de macacos, o que sugere que há um mecanismo cognitivo correspondente que as espécies acima, mas não outras, possuem. No entanto, foi apontado que existem restrições ecológicas e biológicas neste teste: nem todas as espécies são orientadas visualmente e algumas delas acham os olhos aversivos (essa foi a explicação para os estudos que falharam em mostrar a MSR em gorilas). Para uma discussão dessas questões, consulte a coletânea de artigos em *Self-awareness in Animals and Humans* (PARKER et al., 1994).

#### 4.4 Metacognição, Memória, e Monitoramento de Incerteza

A pesquisa sobre metacognição é outra área de pesquisa que busca investigar a compreensão de que alguém tem de seus próprios estados mentais (BERAN et al., 2012; CRYSTAL; FOOTE, 2009; SHETTLEWORTH; SUTTON, 2006). Confira Proust (2013) para uma discussão sobre metacognição em primatas. A metacognição está relacionada ao autoconhecimento e também à consciência (*vide* seção **Autoconsciência e Metacognição**).

Por exemplo, aqueles que sabem o que sabem e o que não sabem demonstram metacognição sobre seus estados epistêmicos. Vários testes não verbais para monitoramento de incerteza foram desenvolvidos para uso com diferentes espécies. O paradigma pode ser o seguinte: os sujeitos são treinados para indicar se um estímulo é igual ou diferente ao de uma amostra. Quando os sujeitos respondem corretamente, eles são recompensados com comida, mas a comida é retirada quando eles dão respostas erradas. Uma vez que os sujeitos são treinados nessa tarefa, o paradigma é modificado para introduzir uma chave de “resgate de isca” com a função de iniciar um novo teste sem fornecer recompensa

ou punição. Intercalados com os estímulos fáceis estão estímulos ambíguos que o sujeito é incapaz de categorizar com precisão acima de um nível de chance. Se os sujeitos aprendem a escolher a chave de “resgate” quando estão incertos, pensa-se que isso indica que os sujeitos estão cientes de seu estado epistêmico. Foi relatado que muitas espécies escolhem a chave de “resgate” de forma a maximizar as recompensas, incluindo golfinhos (SMITHE *et al.*, 1995), macacos-rhesus (HAMPTON, 2001), grandes símios (CALL; CARPENTER, 2001) e bebês humanos (CALL; CARPENTER, 2001). Resultados mistos foram relatados com pombos (SOLE *et al.*, 2003).

O monitoramento da memória também pode estar envolvido em algumas tarefas metacognitivas. O psicólogo Robert Hampton descobriu que os macacos-rhesus sabem se podem ou não se lembrar de ter visto uma imagem. Depois de treinar macacos em uma tarefa simples de comparação atrasada com a amostra, Hampton (2001) permitiu que os macacos decidissem se deveriam ou não fazer o teste. Se eles fizessem o teste e passassem, eles recebiam um prêmio valioso, mas se eles falhassem na tarefa, não recebiam nada. Se, por outro lado, decidissem não fazer o teste, recebiam uma recompensa alimentar de menor valor. Hampton descobriu que a frequência com que o macaco optou por não fazer o teste aumentou com a duração do atraso desde a apresentação da amostra, e que os macacos foram capazes de maximizar suas recompensas ao julgar corretamente quando eles poderiam passar na tarefa. Em um conjunto mais recente de estudos, Hampton e seus alunos testaram sete hipóteses alternativas e concluíram que a melhor explicação para suas descobertas é que os macacos estão monitorando seus estados de memória (BASILE *et al.*, 2015).

Embora esses testes tenham sido projetados para testar a metacognição, Peter Carruthers argumenta que os animais podem resolver os problemas sem se envolver em raciocínios de segunda ordem. Ele sugere que o animal pode estar operando sobre crenças e desejos de diferentes forças, e que sistemas de raciocínio prático padrão podem ser usados para produzir respostas diferentes para as diferentes permutações de crenças e desejos fortes e fracos (CARRUTHERS, 2008). Alternativamente, os animais podem basear suas escolhas em seus estados afetivos, em vez de estados representacionais metacognitivos (CARRUTHERS; RITCHIE, 2012). No entanto, tais respostas podem estar baseadas em concepções distintas da natureza da metacognição. Força de crença e sentimentos afetivos são comumente discutidos na literatura de psicologia como exemplos de metacognição. Por exemplo, psicólogos criaram modelos metacognitivos para mostrar que a força dos traços de



resposta pode ser usada para resolver problemas metacognitivos (SMITH *et al.*, 2008). Além disso, a pesquisa sobre as capacidades metacognitivas humanas estuda o sentimento de correção como um julgamento metacognitivo. Por exemplo, Valerie Thompson descobre que a sensação de acerto de uma resposta inicial prediz tanto o tempo de resposta quanto a probabilidade de mudar a resposta original (THOMPSON *et al.*, 2011).

#### 4.5 Prática Moral

Recentemente, houve um ressurgimento do interesse em examinar se outras espécies compartilham com os humanos qualquer uma das faculdades envolvidas na moralidade ou no engajamento normativo.

A visão de que os animais podem ser agentes morais desenvolvidos é defendida por Bekoff e Pierce (2009), que argumentam que algumas espécies têm uma forma distinta de moralidade que não é uma precursora da moralidade humana. Uma vez que eles consideram 'moralidade' como significando "um conjunto de comportamentos relacionados aos outros que cultivam e regulam interações complexas dentro de grupos sociais" (BEKOFF; PIERCE, 2009, 82), eles assumem que a complexidade do comportamento animal, a organização social e flexibilidade cognitiva demonstram que outras espécies possuem moralidade nesse sentido. Um ponto central é que espécies diferentes possuem normas diferentes, e isso torna a moralidade animal relativa à espécie. Apesar das diferenças, eles afirmam que as semelhanças importantes entre as espécies incluem as capacidades de empatia, altruísmo, cooperação e talvez um senso de equidade [*fairness*]. Se tais afirmações sobre as capacidades dos animais são verdadeiras ou não, é uma questão de muitas pesquisas atuais.

Outros são mais circunspectos. Mark Rowlands (2012) argumenta que, dado que os animais podem ter emoções morais, e que essas emoções morais fornecem razões morais para as ações de um animal, os animais podem se envolver na prática moral sendo sujeitos morais. Sujeitos morais são capazes de agir por razões morais, mas não são agentes morais responsáveis por suas ações, porque não podem considerar essas razões morais, ou agir para mudá-las. Os animais podem agir por razões cujo conteúdo inclui emoções morais; portanto, de acordo com sua visão, podemos buscar evidências da moralidade animal observando se os animais podem, de maneira emocional e confiável, responder às características moralmente salientes de suas situações.

Frans de Waal (1997, 2006, 2009) também assume que a emoção e a empatia desempenham um papel fundamental na moralidade humana, e ele usa as evidências de empatia e reciprocidade em outros animais para indicar que a moralidade evoluiu, de modo que o comportamento moral humano é um desenvolvimento de capacidades mais antigas que vemos em outras espécies. Vemos empatia em outros animais quando eles oferecem ajuda a outro que é diferente do tipo de ajuda que o próprio ator precisa. De Waal dá vários exemplos que são discutidos na próxima seção. A reciprocidade é vista em outros animais quando eles tratam os outros de maneira diferente, dependendo da maneira como são tratados. Ele interpreta sua pesquisa empírica sobre jogos econômicos em chimpanzés (PROCTOR *et al.*, 2013) e sobre equidade [*fairness*] em macacos-prego (DE WAAL; BROSANAN, 2003) como evidência de reciprocidade e empatia em outros primatas.

Além disso, Gary Varner (2012) argumenta que, embora os animais não sejam pessoas inteiramente, alguns animais possuem o status moral de quase-pessoas. Pessoas têm capacidades cognitivas que Varner acha que faltam aos outros animais: um sentido narrativo de eu [*self*], que requer racionalidade, autoconsciência e autonomia no sentido de ter desejos de segunda ordem e a capacidade de pensar sobre si mesmo, sobre seu nascimento, morte e personalidade. Para ter essas capacidades, argumenta Varner, é preciso linguagem. Em vez de pessoas, animais como chimpanzés, golfinhos, elefantes e *scrub-jays* são quase-pessoas; ratos, macacos e papagaios também podem ser quase-pessoas. Quase-pessoas podem engajar-se em pensamentos sobre o passado e futuro, para que possam re-experienciar conscientemente os eventos e fazer planos. Essa capacidade dá às quase-pessoas oportunidades adicionais de felicidade, porque podem reviver experiências agradáveis, e de infelicidade, porque podem temer um futuro desagradável. Além disso, uma quase-pessoa tem um sentido presente de eu [*self*], na medida em que pode ler mentes, e Varner conclui que macacos, elefantes, golfinhos e *scrub-jays* têm a capacidade de atribuir estados perceptivos aos outros.

#### 4.5.1 Emoção e Empatia

A cognição social não se limita a conhecer as razões do outro para agir. Ela também pode incluir a compreensão das emoções dos outros. Entre as espécies sociais, a consciência das emoções dos outros pode desempenhar um papel na regulação das interações sociais, na coordenação de comportamento, na formação

de laços entre mães e bebês, bem como na formação de coalizões de curto prazo e relacionamentos duradouros.

Pelo menos desde *The Expression of Emotion in Man and Animals*, de Darwin, as expressões faciais têm sido de interesse porque podem indicar os estados afetivos dos indivíduos. Pesquisas de campo e de laboratório sugerem que as expressões faciais têm os mesmos tipos de funções para chimpanzés e humanos (VAN HOOFF, 1967, 1972; GOODALL, 1986; PARR, 2001). Assim como Paul Ekman defendeu a universalidade das expressões emocionais entre os humanos de todas as culturas (EKMAN *et al.*, 1969), os pesquisadores animais argumentaram que pelo menos algumas emoções humanas também são encontradas nos chimpanzés, e que as expressões faciais dos chimpanzés são homólogas às expressões faciais humanas em morfologia e função. Por exemplo, van Hoof argumentou que a exibição de dentes chimpanzés é homóloga ao sorriso humano (VAN HOOFF, 1973).

A pesquisa de Lisa Parr demonstra que os chimpanzés, como bebês humanos, são capazes de categorizar as expressões faciais associadas a diferentes respostas emocionais. Usando um paradigma de igualação [*match-to-sample*], Parr e colegas mostraram que os chimpanzés reconhecem pelo menos cinco expressões faciais diferentes: exibição de dentes à mostra, grito, piada, face de brincar, e face de lábios relaxados (PARR *et al.*, 1998). Dada a importância dessas expressões faciais para os chimpanzés, Parr e colegas argumentaram que as expressões faciais são comportamentos importantes para regular as relações sociais (PARR; DE WAAL, 1999; PARR *et al.*, 2000). A pesquisa atual no laboratório de Parr está focada no desenvolvimento do ChimpFACS [*Facial Action Coding System* - Sistema de codificação de ação facial] modelado após o trabalho de Ekman em emoções em expressões faciais humanas. Eles estão usando o ChimpFACS para construir modelos de expressões de chimpanzés a fim de determinar a configuração dos movimentos musculares que os chimpanzés consideram salientes em sua percepção de emoção.

Outra área de estudo em emoção animal se concentra no estresse. Como uma resposta a situações potencialmente perigosas, o estresse é considerado uma emoção adaptativa no curto prazo para humanos e outros animais. O estresse é medido fisiologicamente por meio dos níveis de hormônios glicocorticoides, como o cortisol. Em humanos, os níveis de cortisol estão correlacionados com os níveis de estresse, e os pesquisadores estudaram estressores como a dominância e o ranking de status em várias espécies diferentes (SAPOLSKY, 2005; ABBOTT *et al.*, 2003). Por exemplo, entre babuínos, o estresse na forma de níveis elevados de glicocorticoides foi documentado em mulheres por cerca de um mês após a morte

de um parente próximo; em mães que amamentam quando chega um macho migrante potencialmente infanticida; em fêmeas quando o sistema de ranqueamento feminino está passando instabilidade; e em machos quando o sistema de ranqueamento masculino está instável (CHENEY; SEYFARTH, 2007). No entanto, os babuínos não parecem sentir estresse em face do estresse uns dos outros. As mulheres não estão preocupadas com a instabilidade do ranqueamento masculino, e a morte de um bebê de um bando não aumenta os níveis de estresse de ninguém, exceto da mãe. Sugere-se que, para os babuínos, o estresse é pessoal e egocêntrico, e a falta de sensibilidade ao estresse dos outros pode ser também um indicativo de falta de empatia (CHENEY; SEYFARTH, 2007).

A emoção em outros animais está sendo estudada por meio de técnicas de imagem cerebral, como fMRI e PET. Pesquisadores interessados em PTSD e estresse estudam os cérebros de macacos, ratos e camundongos (MARZLUFF; ANGELL, 2012). Usando PETs, John Marzluff e colegas podem ver que diferentes partes do cérebro do corvo estão ativas quando o corvo vê uma pessoa conhecida perigosa em comparação com uma pessoa conhecida amistosa, e que essas regiões do cérebro são homólogas às regiões do cérebro que são ativas em humanos nos mesmos tipos de situações emocionais. Marzluff conclui que os pássaros sentem medo de uma forma semelhante à forma como os humanos experimentam o medo (MARZLUFF *et al.*, 2012).

Algumas pesquisas em psicologia moral sugerem que a empatia é um componente necessário da agência moral, e os pesquisadores da cognição animal têm examinado se alguma outra espécie compartilha essa habilidade com os humanos. Pensa-se que a empatia requer a mesma sofisticação cognitiva que a compreensão do estado mental ou intenção de outra pessoa, mas, além disso, requer uma resposta afetiva a esse estado mental. Embora os termos **empatia** (compartilhar o estado mental de outra pessoa) e **simpatia** (um sentimento amigável em resposta ao estado mental de outra) tenham significados distintos na história da psicologia, na psicologia do senso comum e na teoria ética, na pesquisa da cognição animal os significados são frequentemente borrados e os termos são usados indistintamente.

Os primeiros relatos de empatia por um chimpanzé vieram do psicólogo comparativo russo N. N. Ladygina-Kohts, que criou um chimpanzé chamado Joni no início do século 20 (LADYGINA-KOHTS, 2002). Mais recentemente, Sanjida O'Connell analisou milhares de relatórios qualitativos de respostas de primatas ao sofrimento de outros, e seus resultados sugerem que os macacos dão respostas

complexas diante das emoções de outros macacos, em comparação com as respostas de macacos em situações semelhantes (O'CONNELL, 1995). Estudos do comportamento de chimpanzés realizados por Frans de Waal e seus colegas sugerem que os chimpanzés entendem as emoções e respondem a diferentes estados emocionais com comportamentos diferentes. Por exemplo, consolar o perdedor de uma luta, ajudar, etc. De Waal considera esses comportamentos como evidência de empatia em chimpanzés (DE WAAL, 2006).

#### 4.5.2 Cooperação

Muitas das pesquisas sobre empatia e altruísmo em outras espécies examinam o comportamento de ajuda. A cooperação e ajuda entre humanos é onipresente, mesmo quando exige que o agente sofra um custo, e mesmo quando o destinatário é um estranho. No entanto, como a ajuda pode ser feita sem empatia, e a ajuda empática requer recursos cognitivos adicionais (por exemplo, conhecimento sobre como ajudar alguém a atingir seu objetivo e a motivação para agir com base nesse conhecimento), é importante entender as limitações dos dados.

Relatos de comportamento animal naturalista sugerem que muitas espécies animais não humanas se engajam em comportamento pró-social que pode ser empático ou protoempático por natureza (DE WAAL, 1996). De Waal frequentemente apresenta o famoso exemplo de Binti-Jua, uma gorila fêmea do zoológico Brookfield em Chicago, que virou notícia ao resgatar um menino humano de 3 anos que caiu em seu habitat. Binti-Jua embalou o menino em seus braços antes de entregá-lo a um tratador. No entanto, os críticos da interpretação pró-social do comportamento de Binti-Jua sugerem que, dada sua exposição inicial à brincadeira com bonecas, uma explicação de aprendizagem associativa também é possível.

Outros afirmam que o que pode parecer comportamento pró-social pode, em vez disso, ser uma forma de eliminar estímulos aversivos. Por exemplo, pesquisas com ratos e macacos-rhesus mostraram que ambas as espécies param de comer quando isso causa choques em outro membro da espécie em uma gaiola adjacente (MASSERMAN *et al.*, 1964). Masserman relata que um macaco-rhesus quase morreu de fome para evitar choques no outro. Alternativamente, o comportamento de ajuda entre parentes pode ser explicado de forma não cognitiva como altruísmo biológico. Para determinar se outras espécies se envolvem em um comportamento de ajuda que não pode ser explicado por outros mecanismos, os pesquisadores

desenvolveram paradigmas para determinar se os chimpanzés exibem um comportamento de ajuda a indivíduos com os quais não estão relacionados. Os chimpanzés são considerados uma espécie especialmente boa para investigar, dada a gama de comportamentos cooperativos que eles desempenham naturalmente, como caça (BOESCH, 2002), patrulhamento de fronteira (MITANI, 2002) e formação de coalizões (DE WALL, 1982). A cooperação entre chimpanzés (HIRATA, 2003; MELIS *et al.*, 2006) e bonobos (HARE *et al.*, 2007) foi demonstrada em uma tarefa de compartilhamento de alimentos, mas se acredita que os chimpanzés cooperem apenas quando as díades são geralmente tolerantes umas com as outras (HARE *et al.*, 2007).

Nos últimos anos, uma série de estudos experimentais investigaram os contextos nos quais os chimpanzés atuam para ajudar os outros, e compararam o comportamento do chimpanzé com o comportamento humano. Diferenças e semelhanças foram encontradas, sugerindo a alguns que as raízes evolutivas do comportamento de ajuda podem ser diferentes em chimpanzés e humanos (BROSAN; BSHARY, 2010; GREENBERG *et al.*, 2010; MELIS; SEMMANN, 2010; YAMAMOTO; TANAKA, 2009). Em um conjunto de estudos, Warneken e colegas compararam o comportamento de ajuda de bebês e chimpanzés humanos em um ambiente social. O experimentador fez pedidos não verbais de ajuda, atingindo assim um objetivo como pegar um objeto caído ou abrir uma caixa. Enquanto crianças de 18 meses e chimpanzés respondem a pedidos simples (por exemplo, pegar um objeto caído), as crianças são capazes de responder a pedidos mais complexos (WARNEKEN; TOMASELLO, 2006). Outros estudos descobriram que os chimpanzés ajudam humanos ou outros chimpanzés a abrir portas para adquirir alimentos (WARNEKEN *et al.*, 2007; MELIS *et al.*, 2008) e respondem a pedidos de outros chimpanzés para trazer uma ferramenta necessária (YAMAMOTO *et al.*, 2009). Os chimpanzés ajudam um membro da mesma espécie a ganhar comida até mesmo depois de já terem sido recompensados, sem a necessidade de um pedido de ajuda (GREENBERG *et al.*, 2010).

Outros experimentos falharam em eliciar qualquer comportamento de ajuda, mesmo quando a ajuda não exigiu nenhum esforço adicional por parte do ator (SILK *et al.*, 2005; JENSEN *et al.*, 2006; VONK *et al.*, 2008). Em um estudo, a configuração experimental permitiu ao agente puxar uma das duas cordas, uma das quais entregava comida apenas para o agente enquanto a outra entregava comida para um chimpanzé adjacente além de para o agente (SILK *et al.*, 2005). Não foi encontrada preferência para puxar a corda que recompensava os dois animais, embora o agente pudesse ver e ouvir o outro chimpanzé. Não houve correlação significativa entre a presença de outro animal e qual corda foi puxada. Silk e colegas (2005, p. 1359)

concluem que “A ausência de preferências relacionadas aos outros em chimpanzés pode indicar que tais preferências são uma propriedade derivada da espécie humana, ligada a capacidades sofisticadas de aprendizagem cultural, teoria da mente, tomada de perspectiva e julgamento moral”.

Várias explicações foram oferecidas para o resultado negativo. Alguns sugerem que os chimpanzés podem ter se distraído com a presença de comida (WARNEKEN; TOMASELLO, 2006), ou que eles são mais propensos a ajudar em resposta a um pedido de ajuda (YAMAMOTO, 2009). No entanto, resultados experimentais adicionais descobriram que chimpanzés ajudarão outros chimpanzés a adquirir comida mesmo quando não há um pedido de assistência. Em vez disso, o design dos estudos pode ter suscitado a natureza competitiva dos chimpanzés (GREENBERG *et al.*, 2010).

A conclusão de que algumas espécies de primatas irão cooperar em alguns contextos levou a uma investigação mais ampla sobre as situações que levam à cooperação. Byrnie e colegas (2015) descobriram que chimpanzés e bonobos tendem a compartilhar alimentos de alto valor mais do que alimentos de baixo valor. Suchak *et al.* (2014) descobriram que os chimpanzés se envolvem espontaneamente em ações conjuntas e, particularmente, preferem cooperar com parentes ou com não parentes do mesmo nível de hierarquia. Molesti e Majolo (2016) descobriram que os macacos selvagens Barbary são mais propensos a cooperar com um parceiro com o qual eles já são conhecidos por afiliar. E, em um grande estudo com 15 espécies de primatas, os pesquisadores descobriram que a reprodução cooperativa é o melhor preditor para o aumento do comportamento pró-social. Eles concluem que a hipercooperação humana que vemos hoje provavelmente evoluiu do movimento hominídeo em direção ao cuidado infantil cooperativo (BURKART *et al.*, 2014). Em resposta a estudos e observações de que grandes símios e outros primatas cooperam em alguns casos, Michael Tomasello insiste que, para os chimpanzés, “a chave para entender sua cooperação é esta mesma matriz abrangente de competição social” (2016, p. 23).

### 4.5.3 Punição

A punição pode ser considerada o outro lado da moeda, em relação à ajuda. Como os comportamentos altruístas, os atos de punição podem diminuir a aptidão individual, e por isso a existência de punição em animais é um quebra-

cabeça para os biólogos. Em muitas espécies, os indivíduos se envolvem em atos agressivos em resposta a violações de seus interesses. No entanto, a análise de teoria dos jogos sugere que a punição é uma estratégia adaptativa para indivíduos que vivem em grupos sociais estáveis, e que a punição pode ajudar a estabelecer ou garantir o domínio em um grupo social (CLUTTON-BROCK; PARKER, 1995).

Estudos experimentais de respostas de primatas a violações de seus interesses (ou de normas de grupo) revelaram resultados mistos. Por exemplo, enquanto os macacos-prego se retraem e param de participar de uma tarefa se observarem outro macaco recebendo uma recompensa melhor pela mesma tarefa (BROSNAN; DE WAAL, 2003), os chimpanzés não rejeitam uma divisão desigual de recursos em um jogo de ultimato (JENSEN *et al.*, 2007). Jensen e seus colegas sugerem que os chimpanzés podem não se preocupar com equidade [*fairness*] e, em vez disso, são maximizadores racionais de recursos. Em resposta, foi argumentado que, assim como o que é considerado justo difere entre as comunidades humanas, o que é considerado injusto [*unfair*] por humanos pode não ser para outras espécies; os pesquisadores devem considerar o comportamento natural a fim de descobrir normas de equidade potenciais (BEKOFF; PIERCE, 2009; BROSNAN; BSHARY, 2010).

Outros estudos testaram diretamente a punição em chimpanzés. Embora haja evidências de que um chimpanzé punirá um ladrão que tenha roubado sua comida (JENSEN *et al.*, 2007), atualmente existem apenas conclusões negativas sobre a questão da punição de terceiros (JENSEN *et al.*, 2007; RIEDL *et al.*, 2010); em um ambiente experimental, nem mesmo as mães retaliarão um indivíduo que rouba comida de seus filhos. No entanto, uma vez que alguns pesquisadores de campo relatam incidentes de punição de terceiros na natureza, mais pesquisas são necessárias.

Os experimentos projetados para examinar as raízes evolutivas das normas sociais também são sugestivos. Os chimpanzés olham por mais tempo para vídeos de infanticídio do que para outros vídeos com estímulos marcantes, mas sem violações de normas (VON ROHR *et al.*, 2010). Se tais descobertas sugerem que elementos normativos estão presentes nas comunidades animais é uma questão que requer mais investigação. Muitas perguntas ainda precisam ser respondidas (e feitas). Os mecanismos de altruísmo, cooperação e punição, a existência de normas sociais, as exigências afetivas do raciocínio moral são todas questões que requerem análise conceitual, bem como investigação empírica do comportamento pró-social em humanos e outros animais.



## 5. Cognição Animal e Filosofia: O que vem a seguir?

Hoje estamos em uma espécie de era de ouro da pesquisa em cognição animal. Diferentes espécies estão sendo estudadas em campo e em laboratórios, e os resultados desses estudos podem ser relevantes para áreas da filosofia, incluindo teoria da ação, agência, crença, conceitos, consciência, cultura, epistemologia, ética, psicologia do senso comum, imagens, linguagem, memória, causalção mental, conteúdo mental, modularidade da mente, percepção, identidade pessoal, razão prática, racionalidade e assim por diante. Parece que todos os dias um novo relatório é publicado, e muitos deles parecem ter implicações teóricas.

É claro que os relatórios científicos devem ser examinados com cuidado para distinguir entre descobertas metodologicamente sólidas e interpretações injustificadas, e nem é preciso dizer que os relatos da mídia popular sobre esses estudos às vezes são enganosos. A epistemologia da pesquisa em cognição animal tem sido uma área com muita atividade recente (*vide* BUCKNER, 2013; CLATTERBUCK, 2015; FITZPATRICK, 2008; HALINA, 2015; MEKETA, 2014; SOBER, 2015; STARZAK, 2016).

A filosofia da cognição animal, como um subcampo da filosofia da ciência, é um lugar onde tais questões metodológicas podem ser examinadas. Para ler mais nesta área, consulte *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*, de Kristin Andrews (2015); o livro de Colin Allen e Marc Bekoff, *Species of Mind: The Philosophy and Biology of Cognitive Ethology*; e a coleção *The Philosophy of Animal Minds*, editado por Robert Lurz. Para uma introdução a uma abordagem psicológica evolucionária para estudar a cognição animal, consulte *Cognition, Evolution, and Behavior* de Sara Shettleworth. Para uma introdução à abordagem etológica para estudar animais, consulte o *Handbook of Ethological Methods*, de Philip Lehner.

### Referência bibliográfica

- ABBOTT, D. *et al.* Are Subordinates Always Stressed? A Comparative Analysis of Rank Differences in Cortisol Levels Among Primates. **Hormones and Behavior**, v. 43, p. 67-82, 2003.
- ALLEN, C. Animal Concepts Revisited: The Use of Self-Monitoring as an Empirical Approach. **Erkenntnis**, v. 51, n. 1, p. 33-40, 1999.
- ALLEN, C. Skeptic's Progress. **Biology and Philosophy**, v. 17, p. 695-702, 2002.

- ALLEN, C. Is Anyone a Cognitive Ethologist? **Biology and Philosophy**, v. 19, p. 589-607, 2004.
- ALLEN, C.; BEKOFF, M. **Species of Mind: The Philosophy and Biology of Cognitive Ethology**. Cambridge MA: MIT Press, 1997.
- ALLEN, C.; HAUSER, M. Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes. *In*: BEKOFF, M.; JAMIESON, D. (ed.). **Readings in Animal Cognition**. Cambridge MA: MIT Press, 1996. p. 47-62.
- ALLEN, C.; SAIDEL, E. The Evolution of Reference. *In*: CUMMINS, D.; ALLEN, C. (ed.). **The Evolution of Mind**. New York: Oxford University Press, 1998. p. 183-202.
- ALLEN-HERMANSON, S. Morgan's Canon Revisited. **Philosophy of Science**, v. 72, p. 608-631, 2005.
- AMY, M. *et al.* Effects of personality on territory defence in communication networks: A playback experiment with radio-tagged great tits. **Proceedings of the Royal Society B: Biological Sciences**, 2010.
- ANDREWS, K. Interpreting Autism: A Critique of Davidson on Thought and Language. **Philosophical Psychology**, v. 15, p. 317-332, 2002.
- ANDREWS, K. Chimpanzee Theory of Mind: Looking in all the Wrong Places? **Mind and Language**, v. 20, p. 521-536, 2005.
- ANDREWS, K. Politics or Metaphysics? On Attributing Psychological Properties to Animals. **Biology and Philosophy**, v. 24, n. 1, p. 51-63, 2009.
- ANDREWS, K. Beyond Anthropomorphism: Attributing Psychological Properties to Animals". *In*: BEAUCHAMP, T.; FREY, R. (ed.). **The Oxford Handbook of Animal Ethics**. Oxford University Press, 2011. p. 469-494.
- ANDREWS, K. **The Animal Mind: The Philosophy of Animal Cognition**. Routledge, 2015.
- ANDREWS, K. **Do Ape Read Minds? Toward a New Folk Psychology**. MIT Press, 2012a.
- ANDREWS, K. Review of Lurz Mindreading Animals. **Notre Dame Philosophical Review**, v. 30, p. 2012b.
- ANDREWS, K. A Role for Folk Psychology in Animal Cognition Research. *In*: BLANK, A. (ed.). **Animals: Basic Philosophical Concepts**. Philosophia: Munich, no prelo.
- ANDREWS, K.; HUSS, B. Anthropomorphism, Anthropectomy, and the Null Hypothesis. **Biology and Philosophy**, v. 29, n. 5, p. 711-729, 2014.

- ANDREWS, K.; RADENOVIC, L. Confronting Language, Representation, and Belief: A Limited Defense of Mental Continuity. *In*: VONK, J.; SHACKELFORD, T. (ed.). **The Oxford Handbook of Comparative Evolutionary Psychology**. Oxford: Oxford University Press, 2012.
- AQUINAS, S. T. **Summa Theologica**. Grand Rapids, Christian Classics, 1981.
- ARBIB, M. The mirror system, imitation, and the evolution of language. *In*: DAUTENHAHN, K.; NEHANIV, C. (ed.). **Imitation in animals and artifacts: Complex adaptive systems**. Cambridge: MIT Press, 2002. p. 229-280.
- ARBIB, M.; LIEBAL, K.; PIKA, S. Primate vocalization, gesture, and the evolution of language. **Current Anthropology**, v. 49, n. 6, p. 1053-1076, 2008.
- ARISTOTLE. **The Metaphysics**. New York: Penguin Classics, 1999.
- ARMSTRONG, D. **Belief, Truth and Knowledge**. Cambridge, MA: Cambridge University Press, 1973.
- ASQUITH, P. Primate Research Groups in Japan: Orientations and East-West Differences. *In*: FEDIGAN, L.; ASQUITH, P. (ed.). **The Monkeys of Arashiyama. Thirty-five Years of Research in Japan and the West**. Albany: SUNY Press, 1991. p. 81-98.
- ASQUITH, P. Why Anthropomorphism is not Metaphor: Crossing Concepts and Cultures in Animal Behavior Studies. *In*: MITCHELL, R.; THOMPSON, N.; MILES, H. (ed.). **Anthropomorphism, Anecdotes, and Animals**. Albany: State University of New York, 1997. p. 22-36.
- BAILLARGEON, R.; DEVOS, J. Object Permanence in Young Infants: Further Evidence. **Child Development**, v. 62, p. 1227-1246, 1991.
- BARLOW, H. The Mechanical Mind. **Annual Review of Neuroscience**, v. 13, p. 15-24, 1990.
- BARRETT, L. **Beyond the Brain: How Body and Environment Shape Animal and Human Minds**. Princeton, NJ: Princeton University Press, 2011.
- BARRETT, L.; HENZI, P. The Social Nature of Primate Cognition. **Proceedings of the Royal Society B: Biological Sciences**, v. 272, p. 1865-1875, 2005.
- BARRETT, L.; HENZI, P.; KENDALL, D. Social brains, simple minds: Does social complexity really require cognitive complexity? **Philosophical Transactions of the Royal Society B**, v. 362, n. 1480, p. 561-575, 2007.
- BARTON, R.; DUNBAR, R. Evolution of the Social Brain. *In*: WHITEN, A.; BYRNE, R. (ed.). **Machiavellian Intelligence II: Evaluations and Extensions**. Cambridge: Cambridge University Press, 1997. p. 240-263.

- BASILE, B. *et al.* Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. **Journal of Experimental Psychology General**, v. 144, n. 1, p. 85-102, 2015.
- BATESON, P.; MAMELI, M. The innate and the acquired: Useful clusters of a residual distinction from folk biology? **Developmental Psychobiology**, v. 49, n. 8, p. 818-831, 2007.
- BECK, B. Chimpoentrism: Bias in Cognitive Ethology. **Journal of Human Evolution**, v. 11, p. 3-17, 1982.
- BECK, J. The Generality Constraint and the Structure of Thought. **Mind**, v. 121, n. 483, p. 563-600, 2012.
- BECK, J. Why We Can't Say What Animals Think. **Philosophical Psychology**, v. 26, n. 4, p. 520-546, 2013.
- BEECHER, M. *et al.* Bird Song Learning in an Eavesdropping Context. **Animal Behavior**, v. 73, p. 929-935, 2007.
- BEKOFF, M. Social Play Behaviour: Cooperation, Fairness, Trust, and the Evolution of Morality. **Journal of Consciousness Studies**, v. 8, p. 81-90, 2001.
- BEKOFF, M.; ALLEN, C. Cognitive Ethology: Slayers, Skeptics and Proponents. *In*: MITCHELL, R.; THOMPSON, N.; MILES, H. (ed.). **Anthropomorphism, Anecdotes, and Animals**. Albany: State University of New York Press, 1997. p. 313-334.
- BEKOFF, M.; PIERCE, J. **Wild Justice: The Moral Lives of Animals**. Chicago: University of Chicago Press, 2009.
- BENNETT, J. Some Remarks about Concepts. **Behavioral and Brain Sciences**, v. 1, p. 557-560, 1978.
- BERAN, M. *et al.* (ed.). *In*: **The Foundations of Metacognition**. Oxford University Press, 2012. p. 76.
- BERMÚDEZ, J. **Thinking Without Words**. Cambridge MA: MIT Press, 2003a.
- BERMÚDEZ, J. Nonconceptual content: From perceptual experience to subpersonal computational states. *In*: GUNTHER, Y. (Ed.). **Essays in Nonconceptual content**. Cambridge, MA: MIT Press, 2003b.
- BERMÚDEZ, J. Animal Reasoning and Proto-Logic. *In*: HURLEY, S.; NUDDS, M. (ed.). **Rational Animals?** Oxford: Oxford University Press, 2006. p. 127-138.
- BERMÚDEZ, J. Mindreading in the animal kingdom. *In*: LURZ, R. (Ed.). **The Philosophy of Animal Minds**. New York: Cambridge University Press, 2009. p. 145-164.

- BEYNON, P.; RASA, O. Do Dwarf Mongooses Have a Language? Warning Vocalizations Transmit Complex Information. **Suid-Afrikaanse Tydskrif vir Wetenskap**, v. 85, p. 447-450, 1989.
- BLUMBERG, M.; WASSERMAN, E. Animal Mind and the Argument from Design. **American Psychologist**, v. 50, p. 133-144, 1995.
- BOESCH, C. Cooperative Hunting Roles among Taie Chimpanzees. **Human Nature**, v. 13, p. 27-46, 2002.
- BOLHUIS, J.; OKANOYA, K.; SCHARFF, C. Twitter evolution: converging mechanisms in birdsong and human speech. **Nature Reviews Neuroscience**, v. 11, n. 11, p. 747-759, 2010.
- BOYSEN, S.; BERNTSON, G. Numerical Competence in a Chimpanzee. **Journal of Comparative Psychology**, v. 103, p. 23-31, 1989.
- BRANDON, R. **Making It Explicit: Reasoning, Representing, and Discursive Commitment**. Cambridge, MA: Harvard University Press, 1994.
- BRAUER, J. *et al.* Making Inference about the Location of Hidden Foods: Social Dog, Causal Ape. **Journal of Comparative Psychology**, v. 120, p. 38-47, 2006.
- BROSNAN, S.; BSHARY, R. Cooperation and deception: from evolution to mechanisms. **Philosophical Transactions of the Royal Society B**, v. 365, p. 2593-2598, 2010.
- BROWN, J. **The Evolution of Behavior**. New York: W.W. Norton, 1975.
- BUCKNER, C. Two approaches to the distinction between cognition and mere association. **International Journal for Comparative Psychology**, v. 24, n. 1, p. 1-35, 2011.
- BUCKNER, C. Morgan's Canon, meet Hume's Dictum: Avoiding anthropofabulation in cross-species comparisons. **Biology & Philosophy**, v. 28, n. 5, p. 853-871, 2013.
- BUGNYAR, T.; REBER, S.; BUCKNER, C. Ravens attribute visual access to unseen competitors. **Nature Communications**, v. 7, p. 10506, 2016.
- BUGNYAR, T. *et al.* Ravens judge competitors through experience with play caching. **Current Biology**, v. 17, n. 20, p. 1804-1808, 2007.
- BURKART, J. *et al.* The evolutionary origin of human hyper-cooperation. **Nature Communications**, v. 5, p. 4747, 2014.
- BYRNE, R. What's the Use of Anecdotes? Distinguishing Psychological Mechanisms in Primate Tactical Deception. *In*: MITCHELL, R.; THOMPSON, N.; MILES, H. (ed.). **Anthropomorphism, Anecdotes, and Animals**. Albany: SUNY Press, 1997.

- BYRNE, R. Imitation Without Intentionality: Using String Parsing to Copy the Organization of Behavior. **Animal Cognition**, v. 2, p. 63-72, 1999.
- BYRNE, R.; RUSSON, A. Learning by Imitation: A Hierarchical Approach. **Behavioral and Brain Sciences**, v. 21, p. 667-721, 1998.
- BYRNE, R.; WHITEN, A. (ed.). **Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans**. New York: Oxford University Press, 1988.
- BYRNIT, J. Primates socio-cognitive abilities: What kind of comparisons makes sense? **Integrative Psychological & Behavioral Science**, v. 49, n. 3, p. 485-511, 2015.
- CALL, J. Inferences about the location of food in the great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus). **Journal of Comparative Psychology**, v. 118, n. 2, p. 232-241, 2004.
- CALL, J. Inferences by exclusion in the great apes: the effect of age and species. **Animal Cognition**, v. 9, n. 4, p. 393-403, 2006.
- CALL, J. *et al.* "Unwilling" versus "Unable": Chimpanzees' Understanding of Intentional Action. **Developmental Science**, v. 7, n. 4, p. 488-498, 2004.
- CALL, J.; CARPENTER, M. Do Apes and Children Know What They Have Seen? **Animal Cognition**, v. 4, p. 207-220, 2001.
- CAMP, E. Thinking with maps. **Philosophical Perspectives**, v. 21, p. 145-182, 2007.
- CAMP, E. A language of baboon thought? *In*: LURZ, R. (Ed.). **Philosophy of Animal Minds**. New York: Cambridge University Press, 2009.
- CANDLAND, D. **Feral Children and Clever Animals**. New York: Oxford University Press, 1993.
- CARRUTHERS, P. On being simple minded. **American Philosophical Quarterly**, v. 41, p. 205-220, 2004.
- CARRUTHERS, P. Meta-cognition in Animals: A Skeptical Look. **Mind and Language**, v. 23, n. 1, p. 58-89, 2008.
- CARRUTHERS, P. Invertebrate concepts confront the generality constraint. *In*: LURZ, R. (ed.). **The Philosophy of Animal Minds**. New York: Cambridge University Press, 2009.
- CARRUTHERS, P.; RITCHIE, J. The Emergence of Metacognition: Affect and Uncertainty in Animals. *In*: BERAN, M. *et al.* (ed.). **The Foundations of Metacognition**. Oxford University Press, 2012. p. 76.
- CAVALIERI, P.; SINGER, P. **The Great Ape Project: Equality Beyond Humanity**. New York: St. Martin's Press, 1994.

- CHATER, N.; HEYES, C. Animal Concepts: Content and Discontent. **Mind and Language**, v. 9, p. 209-246, 1994.
- CHENEY, D. **Baboon Metaphysics: The Evolution of a Social Mind**. Chicago: University of Chicago Press, 2007.
- CHENEY, D.; SEYFARTH, R. **How Monkeys See the World: Inside the Mind of Another Species**. Chicago: The University of Chicago Press, 1990.
- CHOMSKY, N. **Language and Mind**. New York: Harcourt, Brace; World, 1968.
- CHOMSKY, N. Human Language and Other Semiotic Systems. *In*: SEBEOK, T.; UMIKER-SEBEOK, J. (ed.). **Speaking of Apes: A Critical Anthology of Two-Way Communication with Man**. New York: Plenum Press, 1980. p. 429-440.
- CLARK, A. Reasons, robots, and the extended mind (Rationality for the new millenium). **Mind and Language**, v. 16, n. 2, p. 121-145, 2001.
- CLATTERBUCK, H. Chimpanzee Mindreading and the Value of Parsimonious Mental Models. **Mind & Language**, v. 30, n. 4, p. 414-436, 2015.
- CLUTTON-BROCK, T.; A, P. G. Punishment in animal societies. **Nature**, v. 6511, p. 209-216, 1995.
- CORBALLIS, M. On the evolution of language and generativity. **Cognition**, v. 44, n. 3, p. 197-126, 1992.
- CORBALLIS, M. **From Hand to Mouth, the Origins of Language**. Princeton: Princeton University Press, 2002.
- CORBALLIS, M. From Hand to Mouth: The Gestural Origins of Language. *In*: CHRISTIANSEN, M.; KIRBY, S. (ed.). **Language Evolution**. Oxford: Oxford University Press, 2003. p. 201-218.
- COSMIDES, L.; TOOBY, J. Origins of Domain-Specificity: The Evolution of Functional Organization. *In*: HIRSCHFELD, L.; GELMAN, S. (ed.). **Mapping the Mind: Domain-Specificity in Cognition and Culture**. New York: Cambridge University Press, 1994. p. 85-116.
- CRIST, E. **Images of Animals: Anthropomorphism and Animal Mind**. Philadelphia: Temple University Press, 1999.
- CRYSTAL, J.; FOOTE, A. Metacognition in animals. **Comparative Cognition & Behavior Reviews**, v. 4, p. 1-16, 2009.
- DALLY, J.; EMERY, N.; CLAYTON, N. Food-caching Western scrub-jays keep track of who was watching when. **Science**, v. 312, p. 1662-1665, 2006.
- DARWIN, C. **The Expression of the Emotions in Man and Animals**. New York: Appleton; Company, 1886.

- DARWIN, C. **The Descent of Man and Selection in Relation to Sex**. Detroit: Gale Research, 1974.
- DASTON, L.; MITMAN, G. (ed.). **Thinking with Animals: New Perspectives on Anthropomorphism**. New York: Columbia University Press, 2005.
- DAVIDSON, D. Thought and Talk. *In*: GUTTENPLAN, S. (Ed.). **Mind and Language**. Oxford: Oxford University Press, 1975. p. 7-23.
- DAVIDSON, D. Rational Animals. **Dialectica**, v. 36, p. 317-328, 1982.
- DENNETT, D. Beliefs about Beliefs. **Behavioral and Brain Sciences**, v. 4, p. 568-570, 1978.
- DENNETT, D. Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended. **Behavioral and Brain Sciences**, v. 6, p. 343-390, 1983.
- DENNETT, D. **The Intentional Stance**. Cambridge, MA: MIT Press, 1987.
- DENNETT, D. Do Animals Have Beliefs? *In*: ROITBLAT, H.; MEYER, J. (ed.). **Comparative Approaches to Cognitive**. Science: MIT Press, 1995. p. 111-118.
- DESCARTES, R. **A Discourse on the Method**. Oxford: Oxford University Press, 2006.
- DICKINSON, A. What are association formation models. **Learning and Behavior**, v. 37, p. 21-24, 2009.
- DONALD, M. **Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition**. Cambridge, Mass: Harvard University Press, 1991.
- DRETSKE, F. **Explaining Behavior: Reasons in a World of Causes**. Cambridge: MIT Press, 1988.
- DRETSKE, F. Minimal Rationality. *In*: HURLEY, S.; NUDDS, M. (ed.). **Rational Animals?** Oxford: Oxford University Press, 2006. p. 107-116.
- DUGATKIN, L. **Principles of Animal Behavior**. New York: W. W. Norton & Company, 2013.
- DUMMETT, M. **Seas of Language**. Oxford: Oxford University Press, 1993.
- DUNBAR, R. Neocortex Size as a Constraint on Group Size in Primates. **Journal of Human Evolution**, v. 20, p. 469-493, 1992.
- DUNBAR, R. The Social Brain Hypothesis. **Evolutionary Anthropology**, v. 6, p. 178-190, 1998.
- EKMAN, P.; SORENSON, E.; FRIESEN, W. Pan-Cultural Elements in Facial Displays of Emotions. **Science**, v. 164, n. 3875, p. 86-88, 1969.
- EMERY, N. The Eyes Have It: The Neuroethology, Function, and Evolution of Social Gaze. **Neuroscience and Biobehavioral Reviews**, v. 24, p. 581-604, 2000.
- EMERY, N.; CLAYTON, N. Western Scrub-Jays (*Aphelocoma Californica*) Use Cognitive Strategies to Protect Their Caches from Thieving Conspecifics. **Animal Cognition**, v. 7, p. 37-43, 2004.



- ERDŐHEGYI, Á. *et al.* Dog-logic: inferential reasoning in a two-way choice task and its restricted use. **Animal Behaviour**, v. 74, n. 4, p. 725-737, 2007.
- EVANS, C. Referential Signals. *In*: OWINGS, D.; BEECHER, M.; THOMPSON, N. (ed.). **Perspectives in Ethology, Vol.12: Communication**. New York: Plenum, 1997. p. 99-143.
- EVANS, C. Cracking the Code: Communication and Cognition in Birds. *In*: BEKOFF, M.; ALLEN, C.; BURGHARDT, G. (ed.). **The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition**. Cambridge, MA: MIT Press, 2002. p. 315-322.
- EVANS, C. Representational Signaling in Birds. **Biology Letters**, v. 3, p. 8-11, 2007.
- EVANS, C.; EVANS, L. Chicken Calls are Functionally Referential. **Animal Behavior**, v. 58, p. 307-319, 1999.
- EVANS, C.; EVANS, L.; MARLER, P. On the Meaning of Alarm Calls: Functional Reference in an Avian Vocal System. **Animal Behaviour**, v. 46, p. 23-38, 1993.
- EVANS, C.; MACEDONIA, J.; MARLER, P. Effects of Apparent Size and Speed on the Response of Chickens (*Gallus Gallus*) to Computer-Generated Simulations of Aerial Predators. **Animal Behaviour**, v. 46, p. 1-11, 1993.
- EVANS, C.; MARLER, P. Language and Animal Communication: Parallels and Contrasts. *In*: ROITBLAT, H.; MEYER, J. (ed.). **Comparative Approaches to Cognitive Science**. Cambridge MA: MIT Press, 1995. p. 341-382.
- FARAH, M. Neuroethics and the problem of other minds: Implications of neuroscience for the moral status of brain-damaged patients and nonhuman animals. **Neuroethics**, v. 1, p. 9-18, 2008.
- FISHER, J. The Myth of Anthropomorphism. *In*: BEKOFF, M.; JAMIESON, D. (ed.). **Interpretation and Explanation in The Study Of Animal Behavior**. Boulder: Westview Press, 1990. v. 1p. 96-116.
- FISHER, J. Disambiguating Anthropomorphism: An Interdisciplinary Review. *In*: BATESON, P.; KLOPFER, P. (ed.). **Perspectives in Ethology**. New York: Plenum, 1991. v. 9p. 49-85.
- FITZPATRICK, S. Doing away with Morgan's Canon. **Mind and Language**, v. 23, p. 224-246, 2008.
- FITZPATRICK, S. The Primate Mindreading Controversy: A Case Study in Simplicity and Methodology in Animal Psychology. *In*: LURZ, E. R. W. (ed.). **The Philosophy of Animal Minds**. Cambridge: Cambridge University Press, 2009. p. 224-246.

- FLETCHER, L.; CARRUTHERS, P. Behavior-reading vs. mentalizing in animals. *In*: METCALFE, J.; TERRACE, H. (ed.). **Agency and Joint Attention**. [s.l.] Oxford University Press, 2013. p. 82-99.
- FLOMBAUM, J.; SANTOS, L.; R. Rhesus Monkeys Attribute Perceptions to Others. **Current Biology**, v. 15, p. 447-452, 2005.
- FOUTS, R.; FOUTS, D. The Infant Loulis Learns Signs from Cross-Fostered Chimpanzees. *In*: GARDNER, A.; GARDNER, B.; CANTFORT, T. (ed.). **Teaching sign Language to Chimpanzees**. Albany: State University of New York Press, 1989. p. 280-292.
- GALLUP, G. Chimpanzees: Self-Recognition. **Science**, v. 167, p. 341-343, 1970.
- GARDNER, B.; GARDNER, R. Two-Way Communication with an Infant Chimpanzee. *In*: SCHRIER, A.; STOLLNITZ, F. (ed.). **Behavior of Nonhuman Primates**. New York: Academic Press, 1971. p. 117-184.
- GARDNER, R.; GARDNER, B. Comparative Psychology and Language Acquisition. **Annals of the New York Academy of Sciences**, v. 309, p. 37-76, 1978.
- GENTNER, T. et al. Recursive syntactic pattern learning by songbirds. **Nature**, v. 440, n. 7088, p. 1204-1207, 2006.
- GENTRY, E. et al. Gestural communication of the gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. **Animal Cognition**, v. 12, p. 527-547, 2009.
- GERGELY, G.; NADASDY, Z. Taking the Intentional Stance at 12 Months of Age. **Cognition**, v. 56, p. 165-193, 1995.
- GLOCK, H. Animals, thoughts, and concepts. **Synthese**, v. 123, p. 35-64, 2000.
- GLOCK, H. Can animals act for reasons? **Inquiry**, v. 52, p. 232-254, 2009.
- GLOCK, H. Can animals judge? **Dialectica**, v. 64, p. 11-33, 2010.
- GOLANI, I. A Mobility Gradient in the Organization of Movement: The Perception of Movement through Symbolic Language. **Behavioral and Brain Sciences**, v. 15, p. 249-308, 1992.
- GOLDMAN, A. **Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading**. Oxford: Oxford University Press, 2006.
- GÓMEZ, J. Species Comparative Studies and Cognitive Development. **Trends in Cognitive Science**, v. 9, n. 3, p. 118-125, 2005.
- GÓMEZ, J. Pointing behaviours in apes and human infants: A balanced interpretation. **Child Development**, v. 78, n. 3, p. 729-734, 2007.
- GOODALL, J. **The Chimpanzees of Gombe: Patterns of Behavior**. Cambridge, MA: Harvard University Press, 1986.

- GREENE, E.; MEAGHER, T. Red Squirrels, *Tamiasciurus Hudsonicus*, Produce Predator-Specific Alarm Calls. **Animal Behavior**, v. 55, p. 511-518, 1998.
- GREENFIELD, P. Language, Tools, and Brain: The Ontogeny and Phylogeny of Hierarchically Organized Sequential Behavior. **Behavioral and Brain Sciences**, v. 14, p. 531-595, 1991.
- GRICE, H. Utterer's Meaning and Intention. **The Philosophical Review**, v. 78, p. 147-177, 1969.
- GRIFFIN, D. **Animal Thinking**. Harvard University Press, 1984.
- GRIFFIN, D. Animal Consciousness. **Neuroscience and Biobehavioral Reviews**, v. 9, p. 615-22, 1985.
- GRIFFIN, D. **Animal Minds**. Chicago: University of Chicago Press, 1992.
- GRÜTER, C.; FARINA, W. The honeybee waggle dance: Can we follow the steps? **Trends in Ecology and Evolution**, v. 24, p. 242-247, 2009.
- GRÜTER, C.; M.S., B.; M, F. W. Informational conflicts created by the waggle dance. **Proceedings of the Royal Society B: Biological Sciences**, v. 275, p. 1321-1327, 2008.
- HAESLER, S. et al. Incomplete and Inaccurate Vocal Imitation after Knockdown of FoxP2 in Songbird Basal Ganglia Nucleus Area X. **PLoS Biology**, v. 5, n. 12, p. 2885-2897, 2007.
- HAMPTON, R. Rhesus Monkeys Know When They Remember. **Proceedings of the National Academy of Sciences of the United States of America**, 2001.
- HARE, B. Do Chimpanzees Know What Conspecifics Know? **Animal Behavior**, v. 6, n. 1, p. 139-151, 2001.
- HARE, B. Chimpanzees Deceive a Human by Hiding. **Cognition**, v. 101, p. 495-514, 2006.
- HARE, B. et al. The Domestication of Social Cognition in Dogs. **Science**, v. 298, p. 1634-1636, 2002.
- HARE, B. et al. Chimpanzees Know what Conspecifics Do and Do Not See. **Animal Behavior**, v. 59, p. 771-785, 2000.
- HARE, B.; CALL, J.; TOMASELLO, M. Communication of Food Location Between Human and Dog (*Canis Familiaris*). **Evolution of Communication**, v. 2, p. 137-159, 1998.
- HARE, B. et al. Tolerance Allows Bonobos to Outperform Chimpanzees on a Cooperative Task. **Current Biology**, v. 17, p. 619-623, 2007.
- HARE, B.; TOMASELLO, M. Domestic Dogs (*Canis Familiaris*) use Human and Conspecific Social Cues to Locate Hidden Food. **Journal of Comparative Psychology**, v. 113, p. 1-5, 1999.

- HARE, B.; WOODS, V. **The Genius of Dogs: How Dogs Are Smarter than You Think**. Dutton Adult, 2013.
- HARMAN, G. Studying the Chimpanzees' Theory of Mind. **Behavioral and Brain Sciences**, v. 1, p. 576-577, 1978.
- HAUSER, M. Our Chimpanzee Mind. **Nature**, v. 437, p. 60-63, 2005.
- HAUSER, M. **Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong**. New York: Ecco, 2006.
- HAUSER, M.; CAREY, S. Building a Cognitive Creature from a Set of Primitives: Evolutionary and Developmental Insights. In: CUMMINS, D.; ALLEN, C. (ed.). **The Evolution of Mind**. New York: Oxford University Press, 1998. p. 51-106.
- HAUSER, M.; CHOMSKY, N.; FITCH, W. The Faculty of Language: What Is It, Who Has It, And How Did It Evolve? **Science**, v. 298, p. 1569-1579, 2002.
- HAUSER, M.; MACNEILAGE, P.; WARE, M. Numerical Representations in Primates. **Proceeding of the National Academy of the Sciences**, v. 93, p. 1514-1517, 1996.
- HAYES, C. **The Ape in Our House**. New York, Harper, 1951.
- HAYES, K.; HAYES, C. Imitation in a Home-Raised Chimpanzee. **Journal of Comparative Psychology**, v. 45, p. 450-459, 1952.
- HEALY, S.; BRAITHWAITE, V. Cognitive ecology: A field of substance? **Trends in Ecology and Evolution**, v. 15, p. 22-26, 2000.
- HERMAN, L. Cognition and Language Competencies of Bottlenosed Dolphins. In: SCHUSTERMAN, R.; THOMAS, J.; WOOD, F. (ed.). **Dolphin Cognition and Behavior: A Comparative Approach**. Hillsdale, NJ: Lawrence Erlbaum Associates, 1986. p. 221-251.
- HERMAN, L. et al. Dolphins (*Tursiops Truncatus*) Comprehend the Referential Character of the Human Pointing Gesture. **Journal of Comparative Psychology**, v. 113, n. 4, p. 347-364, 1999.
- HERRNSTEIN, R. Acquisition, Generalization, and Discrimination Reversal of a Natural Concept. **Journal of Experimental Psychology: Animal Behavior Processes**, v. 5, p. 116-129, 1979.
- HERRNSTEIN, R.; LOVELAND, D.; CABLE, C. Natural Concepts in Pigeons. **Journal of Experimental Psychology: Animal Behavior Processes**, v. 2, p. 285-302, 1976.
- HEYES, C. Theory of Mind in Nonhuman Primates. **Behavioral and Brain Sciences**, v. 21, p. 101-134, 1998.

- HILLIX, W.; RUMBAUGH, D. **Animal Bodies, Human Minds: Ape, Dolphin, and Parrot Language Skills**. New York: Kluwer, 2004.
- HINDE, R. **Animal Behavior: A Synthesis of Ethology and Comparative Psychology**. New York: McGraw-Hill, 1970.
- HIRATA, S. Cooperation in Chimpanzees. **Hattatsu**, v. 95, p. 103-111, 2003.
- HOCKETT, C. **The View From Language: Selected Essays**. Athens, GA: University of Georgia Press, 1977.
- HOLEKAMP, K. Questioning the Social Intelligence Hypothesis. **Trends in Cognitive Science**, v. 11, p. 65-69, 2006.
- HOLLÉN, L.; MANSER, M. Motivation Before Meaning: Motivational Information Encoded in Meerkat Alarm Calls Develops Earlier than Referential Information. **The American Naturalist**, v. 169, p. 758-767, 2007.
- HOPPER, L. et al. Experimental Studies of Traditions and Underlying Transmission Processes in Chimpanzees. **Animal Behaviour**, v. 73, p. 1021-1032, 2007.
- HORNER, V.; WHITEN, A. Causal Knowledge and Imitation/Emulation Switching in Chimpanzees and Children. **Animal Cognition**, v. 8, p. 164-181, 2005.
- HUME, D. **A Treatise of Human Nature**. Oxford: Oxford University Press, 1968.
- HUMPHREY, N. Nature's Psychologists. **New Scientist**, p. 900-904, 1978.
- HURLEY, S. Making Sense of Animals: Interpretation vs. Architecture. **Mind and Language**, v. 18, p. 231-256, 2003a.
- HURLEY, S. Animal Action in the Space of Reasons. **Mind and Language**, v. 18, p. 231-256, 2003b.
- HURLEY, S. Making Sense of Animals. *In*: HURLEY, S.; NUDDS, M. (ed.). **Rational Animals?** Oxford: Oxford University Press, 2006. p. 139-174.
- HURLEY, S.; CHATER, N. (ed.). **Perspectives on Imitation From Neuroscience to Social Science- Volume 1: Mechanisms of Imitation and Imitation in Animals**. Cambridge MA: MIT Press, 2005.
- HURLEY, S.; NUDDS, M. (ed.). **Rational Animals?** Oxford: Oxford University Press, 2006.
- HUTCHINS, E. The role of cultural practices in the emergence of modern human intelligence. **Philosophical Transactions of the Royal Society B**, v. 363, n. 1499, p. 2011-2019, 2008.
- INOUE, S.; MATSUZAWA, T. Working Memory of Numerals in Chimpanzees. **Current Biology**, v. 17, p. 1004-1005, 2007.
- JAMIESON, D. Science, Knowledge, and Animal Minds. **Proceedings of the Aristotelian Society**, v. 98, p. 79-102, 1998.

- JAMIESON, D. What do animals think? *In*: LURZ, R. (ed.). **The Philosophy of Animal Minds**. New York: Cambridge University Press, 2009.
- JANIK, V.; SAYIGH, L.; WELLS, R. Signature Whistle Shape Conveys Identity Information to Bottlenose Dolphins. **Proceedings of the National Academy of Sciences**, v. 103, p. 8293-8297, 2006.
- JOLLY, A. Lemur Social Behavior and Primate Intelligence. **Science**, v. 153, p. 501-506, 1966.
- KANT, I. Duties to Animals and Spirits. **Lectures on Ethics**. New York: Harper & Row, 1963. p. 239-241.
- KARG, K. et al. The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? **Animal Behaviour**, v. 105, p. 211-221, 2015.
- KEELEY, B. Anthropomorphism, Primatomorphism, Mammalomorphism. **Philosophy and Biology**, v. 19, p. 521-540, 2004.
- KEETON, W. **Biological Science**. New York: W.W. Norton, 1967.
- KELLOGG, W.; KELLOGG, L. **The Ape and the Child: A Comparative Study of the Environmental Influence Upon Early Behavior**. New York: Hafner Publishing Company, 1933.
- KENNEDY, J. **The New Anthropomorphism**. Cambridge: Cambridge University Press, 1992.
- KING, B. **The Dynamic Dance: Nonvocal Communication in African Great Apes**. Cambridge, MA: Harvard University Press, 2004.
- KING, B.; SHANKER, S. How can we know the dancer from the dance? The dynamic nature of African great ape social communication. **Anthropological Theory**, v. 3, n. 1, p. 5-26, 2003.
- KIRIAZIS, J. Perceptual Specificity in the Alarm Calls of Gunnison's Prairie Dogs. **Behavioural Processes**, v. 73, p. 29-35.
- KÖHLER, W. **The Mentality of Apes**. New York: Harcourt, Brace, 1925.
- KREBS, J.; DAVIES, N. **An Introduction to Behavioral Ecology**. Oxford: Blackwell Scientific, 1993.
- KRUTZEN, M. et al. Cultural Transmission of Tool Use in Bottlenose Dolphins. **Proceedings of the National Academy of Sciences**, v. 102, p. 8939-8943, 2005.
- LADYGINA-KOHTS, N. **Infant Chimpanzee and Human Child: A Classic 1935 Comparative Study of Ape Emotions and Intelligence**. New York: Oxford University Press, 2002.

- LEAVENS, D.; HOPKINS, W. Intentional Communication by Chimpanzees: A Cross-Sectional Study of the Use of Referential Gestures. **Developmental Psychology**, v. 34, n. 5, p. 813-822, 1998.
- LEAVENS, D.; HOPKINS, W.; BARD, K. Understanding the Point of Chimpanzee Pointing Epigenesis and Ecological Validity. **Current Directions in Psychological Science**, v. 14, n. 4, p. 185-189, 2005.
- LEHNER, P. **Handbook of Ethological Methods**. Cambridge: Cambridge University Press, 1996.
- LIEBAL, K. *et al.* To Move or Not to Move: How Apes Adjust to the Attentional State of Others. **Interaction Studies**, v. 5, n. 2, p. 199-219, 2004.
- LIN, A.; BARD, K.; ANDERSON, J. Development of Self-Recognition in Chimpanzees. **Journal of Comparative Psychology**, v. 106, p. 120-127, 1992.
- LLOYD, E. Kanzi, Evolution, and Language. **Biology and Philosophy**, v. 19, p. 577-588, 2004.
- LOCKE, J. **An Essay Concerning Human Understanding**. Pomona, Italy: Pomona Press, 2007.
- LURZ, R. **The Philosophy of Animal Minds**. New York: Cambridge University Press, 2009.
- LURZ, R. Belief attribution in animals: On how to move forward conceptually and empirically. **Review of Philosophy and Psychology**, v. 2, n. 1, p. 19-59, 2011.
- MAMELI, M.; BATESON, P. An evaluation of the concept of innateness. **Philosophical Transactions of the Royal Society B**, v. 366, n. 1563, p. 436-443, 2011.
- MANSER, M. The Acoustic Structure of Suricates' Alarm Calls varies with Predator Type and the Level of Response Urgency. **Proceedings of the Royal Society B: Biological Sciences**, v. 268, p. 2315-2324, 2001.
- MANSER, M.; BELL, M.; FLETCHER, I. The Information that Receivers Extract from Alarm Calls in Suricates. **Proceedings of the Royal Society B: Biological Sciences**, v. 268, p. 2485-2491, 2001.
- MANSER, M.; R.M., S.; L, C. D. Suricate alarm calls signal predator class and urgency. **Trends in Cognitive Science**, v. 6, n. 2, p. 55-57, 2002.
- MARINO, L.; REISS, D.; GALLUP, G. Mirror Self-Recognition in Bottlenosed Dolphins: Implications for Comparative Investigations of Highly Dissimilar Species. *In*: PARKER, S.; MITCHELL, R.; BOCCIA, M. (ed.). **Self-Awareness in Animals and Humans**. Cambridge: Cambridge University Press, 1994. p. 380-391.

- MARLER, P. Birdsong and Speech Development: Could there be Parallels? **American Scientist**, v. 58, p. 669-673, 1970.
- MARLER, P.; EVANS, C.; HAUSER, M. Animal Signals: Motivational, Referential, or Both? *In*: PAPOUSEK, H.; JÜRGENS, U.; PAPOUSEK, M. (ed.). **Nonverbal Vocal Communication: Comparative and Developmental Approaches**. Cambridge: Cambridge University Press, 1992. p. 66-86.
- MARSH, H.; MACDONALD, S. Orangutans (*Pongo abelii*) "play the odds": Information-seeking strategies in relation to cost, risk, and benefit. **Journal of Comparative Psychology**, v. 126, n. 3, p. 263-278, 2012.
- MARTIN, A.; SANTOS, L. The origins of belief representation: Monkeys fail to automatically represent others' beliefs. **Cognition**, v. 130, n. 3, p. 300-308, 2014.
- MARZLUFF, J.; ANGELL, T. **The Gifts of the Crow: How perception, memory, and emotion enable smart birds to act like humans**. New York: Free Press, 2012.
- MARZLUFF, J. *et al.* Brain imaging reveals neuronal circuitry underlying the crow's perception of human faces. **Proceedings of the National Academy of Sciences**, v. 109, p. 15912-15917, 2012.
- MASSERMAN, J.; WECHKIN, M.; TERRIS, W. Altruistic Behavior in Rhesus Monkeys. **American Journal of Psychiatry**, v. 121, p. 584-85, 1964.
- MATSUZAWA, T. Use of Numbers by a Chimpanzee. **Nature**, v. 315, p. 57-59, 1985.
- MATSUZAWA, T. Koshima Monkeys and Bossou Chimpanzees: Culture in Nonhuman Primates Based on Long-Term Research. *In*: WAAL, F.; TYACK, P. (ed.). **Animal Social Complexity: Intelligence, Culture, and Individualized Societies**. Cambridge, MA: Harvard University Press, 2003.
- MATSUZAWA, T. Sociocognitive Development in Chimpanzees: A Synthesis of Laboratory Work and Fieldwork. *In*: MATSUZAWA, T.; TOMONAGA, M.; TANAKA, M. (ed.). **Cognitive Development in Chimpanzees**. Tokyo: Springer, 2006a. p. 3-36.
- MATSUZAWA, T. Evolutionary Origins of the Human Mother-Infant Relationship. *In*: MATSUZAWA, T.; TOMONAGA, M.; TANAKA, M. (ed.). **Cognitive Development in Chimpanzees**. Tokyo: Springer, 2006b. p. 127-141.
- MCGREW, W. *et al.* Intergroup Differences in a Social Custom of Wild Chimpanzees: The Grooming Hand-Clasp of the Mahale Mountains. **Current Anthropology**, v. 42, p. 148-153, 2001.
- MCGREW, W.; TUTIN, C. Evidence for a Social Custom in Wild Chimpanzees? **Man**, v. 12, p. 234-251, 1978.



- MEKETA, I. A Critique of the Principle of Cognitive Simplicity in Comparative Cognition. **Biology & Philosophy**, v. 29, n. 5, p. 731-745, 2014.
- MELIS, A.; CALL, J.; TOMASELLO, M. Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. **Journal of Comparative Psychology**, v. 120, p. 154-162, 2006.
- MELIS, A.; D, S. How is human cooperation different? **Philosophical Transactions of the Royal Society B**, v. 365, n. 1553, p. 2663-2674, 2010.
- MELIS, A.; HARE, B.; TOMASELLO, M. Chimpanzees Recruit The Best Collaborators. **Science**, v. 311, p. 1297-1300, 2006.
- MENZEL, E. Naturalistic and Experimental Approaches to Primate Behavior. *In*: WILLEMS, E.; RAUSCH, H. (ed.). **Naturalistic Viewpoints in Psychological Research**. New York: Holt, Rinehart; Winston, 1969. p. 78-121.
- MIKLÓSI; TOPAL, J. What Can Dogs Teach Us? **Animal Behavior**, v. 67, p. 995-1004, 2004.
- MILES, H. Apes and Language: The Search for Communicative Competence. *In*: LUCE, J.; WILDNER, H. (ed.). **Language in Primates: Perspectives and Implications**. New York: Springer-Verlag, 1983. p. 43-61.
- MILLIKAN, R. **White Queen Psychology and Other Essays for Alice**. Cambridge, MA: MIT Press, 1993.
- MILLIKAN, R. **Varieties of Meaning: The Jean Nicod Lectures 2002**. Cambridge, MA: MIT Press, 2004.
- MILLIKAN, R. Styles of Rationality. *In*: HURLEY, S.; NUDDS, M. (ed.). **Rational Animals?** Oxford: Oxford University Press, 2006. p. 117-126.
- MITANI, J.; WATTS, D.; MULLER, M. Recent Developments in the Study of Wild Chimpanzee Behavior. **Evolutionary Anthropology**, v. 11, p. 9-25, 2002.
- MITCHELL, C.; DE HOUWER, J.; LOVIBOND, P. The propositional nature of human associative learning. **The Behavioral and Brain Sciences**, v. 32, n. 2, p. 183-198, 2009.
- MITCHELL, R. Anthropomorphic Anecdotalism as Method. *In*: MITCHELL, R.; THOMPSON, N.; MILES, H. (ed.). **Anthropomorphism, Anecdotes, and Animals**. Albany: SUNY Press, 1997. p. 151-169.
- MITCHELL, R.; THOMPSON, N.; MILES, H. **Anthropomorphism, Anecdotes, and Animals**. Albany: State University of New York Press, 1997.
- MOLESTI, S.; MAJOLO, B. Cooperation in wild Barbary macaques: factors affecting free partner choice. **Animal Cognition**, v. 19, n. 1, p. 133-146, 2015.

- MOORE. Ontogenetic constraints on Grice's communication theory. *In*: MATTHEWS, D.; BENJAMINS, J. (ed.). **Developmental Pragmatics in First Language Acquisition**, 2014.
- MOORE. Meaning and Ostension in Great Ape Gestural Communication. **Animal Cognition**, v. 19, p. 223-231, 2016. Publicado online em 2015 antes da versão impressa.
- MORGAN, C. **An Introduction to Comparative Psychology**. London: W. Scott, limited, 1894.
- MORGAN, C. Autobiography. *In*: MURCHISON, C.; BORING, E. G. (ed.). **A History of Psychology in Autobiography**. Clark University Press, 1930.
- NEWEN, A.; BARTELS, A. Animal minds and the possession of concepts. **Philosophical Psychology**, v. 20, n. 3, p. 283-308, 2007.
- NICHOLS, S.; STICH, S. **Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds**. Oxford: UK, Oxford University Press, 2003.
- NISHIDA, T. **Chimpanzees of the Mahale Mountains: Sexual and Life History Strategies**. Tokyo: University of Tokyo Press, 1990.
- O'CONNELL, S. Empathy in Chimpanzees: Evidence for Theory of Mind? **Primates**, v. 36, n. 3, p. 397-410, 1995.
- OWINGS, D.; HENNESSY, D. The Importance of Variation in Sciurid Visual and Vocal Communication. *In*: MURIE, J.; MICHENER, G. (ed.). **The Biology of Ground Dwelling Squirrels**. Lincoln: University of Nebraska Press, 1984. p. 171-200.
- PARKER, S.; BOCCIA, M.; MITCHELL, R. (ed.). **Self-awareness in Animals and Humans: Developmental Perspectives**. Cambridge: Cambridge University Press, 1994.
- PARR, L. Cognitive and Physiological Markers of Emotional Awareness in Chimpanzees (Pan Troglodytes). **Animal Cognition**, v. 4, p. 223-229, 2001.
- PARR, L.; HOPKINS, W.; WAAL, F. The Perception of Facial Expressions in Chimpanzees (Pan Troglodytes). **Evolution of Communication**, v. 2, p. 1-23, 1998.
- PARR, L.; WAAL, F. Visual Kin Recognition in Chimpanzees. **Nature**, v. 399, p. 647-648, 1999.
- PARR, L. et al. Recognizing Facial Cues: Individual Recognition in Chimpanzees and Rhesus Monkeys. **Journal of Comparative Psychology**, v. 114, p. 47-60, 2000.

- PATTERSON, F. The Gestures of a Gorilla: Language Acquisition in Another Pongid. **Brain and Language**, v. 5, p. 72-97, 1978.
- PEARCE, J. **Animal Learning and Cognition**. Hove: Psychology Press, 1997.
- PENN, D. How folk psychology ruined comparative psychology: And how scrub jays can save it. *In*: MENZEL, R.; FISCHER, J. (ed.). **Animal Thinking: Contemporary Issues in Comparative Cognition**. Cambridge: MIT Press, 2011. p. 253-266.
- PENN, D.; K., H. J.; J, P. D. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. **Behavioral and Brain Sciences**, v. 31, p. 109-178, 2008.
- PEPPERBERG, I. **The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots**. Cambridge MA: Harvard University Press, 1999.
- PERRY, S. What Cultural Primatology can Tell Anthropologists about the Evolution of Culture. **Annual Review of Anthropology**, v. 35, p. 171-190, 2006.
- PERRY, S. et al. Social Conventions in Wild White-Faced Capuchin Monkeys. **Current Anthropology**, v. 44, p. 241-258, 2003.
- PETRIDES, M.; CADORET, G.; MACKEY, S. Orofacial Somatomotor Responses in the Macaque Monkey Homologue of Broca's Area. **Nature**, v. 435, n. 30, p. 1235-1238, 2005.
- PFUNGST, O. **Clever Hans (The Horse of Mr. Von Osten)**. New York: Holt, Rinehart; Winston, 1965.
- PIKA, S.; BUGNYAR, T. The Use of Referential Gestures in Ravens (*Corvus Corax*) in the Wild. **Nature Communications**, v. 2, p. 560, 2011.
- PIKA, S.; LIEBAL, K.; CALL, J. Gestural communication of apes. **Gesture**, v. 5, n. 1-2, p. 41-56, 2005.
- PIKA, S.; MITANI, J. Referential gesturing in wild chimpanzees (*Pan troglodytes*). **Current Biology**, v. 16, p. 191-192, 2006.
- PILLEY, J.; REID, A. Border collie comprehends object names as verbal referents. **Behavioral Processes**, v. 86, p. 184-195, 2011.
- PINKER, S. **The Language Instinct**. New York: W. Morrow; Co, 1994.
- PINKER, S.; BLOOM, P. Natural Language and Natural Selection. **Behavioral and Brain Sciences**, v. 13, p. 707-784, 1990.
- PLATT, M. L.; GHAZANFAR, A. A. (ed.). **Primate Neuroethology**. New York: Oxford University Press, 2010.
- PLOOIJ, F. Some Basic Traits of Language in Wild Chimpanzees? *In*: LOCK, A. (ed.). **Action, Gesture and Symbol**. London: Academic Press, 1978. p. 111-131.

- PLOTNIK, J.; WAAL, F.; REISS, D. Self-Recognition in an Asian Elephant. **Proceedings of the National Academy of Sciences**, v. 103, p. 17053-17057, 2006.
- POLLICK, A.; WAAL, F. Ape gestures and language evolution. **Proceedings of the National Academy of Sciences**, v. 104, n. 19, p. 8184-8189, 2007.
- POOLE, J.; GRANLI, P. Signals, Gestures, and Behavior of African Elephants. *In: The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Mammal*. Chicago, Illinois: University of Chicago Press, 2011. p. 109-127.
- POVINELLI, D. **Folk Physics for Apes**. Oxford: Oxford University Press, 2000.
- POVINELLI, D.; DUNPHY-LELII, S. Do chimpanzees seek explanations? Preliminary comparative investigations. **Canadian Journal of Experimental Psychology**, v. 55, p. 185-193, 2001.
- POVINELLI, D.; EDDY, T. What Young Chimpanzees Know About Seeing. **Monographs of the Society for Research on Child Development**, v. 61, n. 247, p. 1-189, 1996.
- POVINELLI, D.; M., B. J.; S, G. Toward a science of other minds: Escaping the argument by analogy. **Cognitive Science**, v. 24, n. 3, p. 509-541, 2000.
- POVINELLI, D.; VONK, J. We Don't Need a Microscope to Explore the Chimpanzee Mind. **Mind and Language**, v. 19, p. 1-28, 2004.
- PREMACK, D. On the Assessment of Language Competence in the Chimpanzee. **Behavior of Nonhuman Primates**, v. 4, p. 185-228, 1971.
- PREMACK, D. Is Language the Key to Human Intelligence. **Science**, v. 303, p. 318-320, 2004.
- PREMACK, D.; PREMACK, A. **Original Intelligence: Unlocking the Mystery of Who We Are**. New York: McGraw-Hill, 2003.
- PREMACK, D.; WOODRUFF, G. Does the Chimpanzee have a Theory of Mind? **Behavioral and Brain Sciences**, v. 1, p. 515-526, 1978.
- PREUTZ, J.; BERTOLANI, P. Savanna chimpanzees, Pan troglodytes verus, hunt with tools. **Current Biology**, v. 17, n. 5, p. 412-7, 2007.
- PRIOR, H.; SCHWARZ, A.; GÜNTÜRKÜN, O. Mirror-induced behavior in the magpie (*Pica pica*): Evidence of self-recognition. **PLoS Biology**, v. 6, n. 8, p. 202, 2008.
- PROCTOR, D. *et al.* Chimpanzees play the ultimatum game. **Proceedings of the National Academy of Sciences**, v. 110, n. 6, p. 2070-2075, 2013.
- PROUST, J. Mind, Space and Objectivity in Non-Human Animals. **Erkenntnis**, v. 51, p. 41-58, 1999.
- PROUST, J. Rationality and Metacognition in Non-Human Animals. *In: HURLEY, S.; NUDDS, M. (ed.). Rational Animals?* Oxford: Oxford University Press,

2006. p. 247-274.
- PROUST, J. **The Philosophy of Metacognition: Mental Agency and Self-Awareness**. Oxford: Oxford University Press, 2013.
- RAJALA, A. *et al.* Rhesus monkeys (*Macaca mulatta*) do recognize themselves in the mirror: Implications for the evolution of self-recognition. **PLoS One**, v. 5, n. 9, p. 12865, 2010.
- RAMSEY, G.; BASTIAN, M.; SCHAIK, C. Animal Innovation Defined and Operationalized. **Behavioral and Brain Sciences**, v. 30, n. 4, p. 393-406, 2007.
- READER, S.; LALAND, K. Animal Innovation: An Introduction. *In*: READER, S.; LALAND, K. (ed.). **Animal Innovation**. New York: Oxford University Press, 2003. p. 3-35.
- REAL, L. Towards a cognitive ecology. **Trends in Ecology and Evolution**, v. 8, p. 413-417, 1993.
- REISS, D. **The Dolphin in the Mirror: Exploring Dolphin Minds and Saving Dolphin Lives**. Boston: Mariner Books, 2012.
- REISS, D.; MARINO, L. Mirror Self-Recognition in the Bottlenose Dolphin: A Case of Cognitive Convergence. **Proceedings of the National Academy of Sciences**, v. 98, p. 5937-5942, 2001.
- RESCORLA, M. Chrysippus' dog as a case study in non-linguistic cognition. *In*: LURZ, R. (ed.). **The Philosophy of Animal Minds**. New York: Cambridge, 2009.
- RESCORLA, R. Pavlovian conditioning: It's not what you think it is. **Journal of Experimental Psychology: Animal Behavior Processes**, v. 43, n. 151-160, 1988.
- RIVAS, J.; BURGHARDT, G. Crotalomorphic: A Metaphor for Understanding Anthropomorphism by Omission. *In*: BEKOFF, M.; ALLEN, C.; BURGHARDT, G. (ed.). **The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition**. Cambridge, MA: MIT Press, 2002. p. 9-18.
- RIZZOLATTI, G.; CRAIGHERO, L. The Mirror-neuron System. **Annual Review of Neuroscience**, v. 27, p. 169-192, 2004.
- R.M., S.; CHENEY, D. The evolution of concepts about agents: Or what do animals recognize when they recognize an agent? *In*: **The Conceptual Mind: New Directions in the Study of Concepts**. Cambridge, MA: MIT Press, 2015.
- ROBERTS, W. **Principles of Animal Cognition**. McGraw-Hill, 1997.
- ROMANES, G. **Animal Intelligence**. Farnborough, England: Gregg International, 1970.
- ROTH, G.; DICKE, U. Evolution of the Brain and Intelligence. **Trends in Cognitive Science**, v. 9, n. 5, p. 250-257, 2005.

- ROWLANDS, M. **Can Animals Be Moral?** Oxford: Oxford University Press, 2012.
- RUIZ, A. *et al.* Do rhesus macaques reason about false beliefs? Artigo apresentado na **International Primatological Association**, 2010.
- RUMBAUGH, D.; GILL, T. Reading and Sentence Completion by a Chimpanzee. **Science**, v. 182, p. 731-733, 1973.
- RUSSON, A. Pantomime in great apes: evidence and implications. **Communicative & Integrative Biology**, no prelo.
- RUSSON, A.; ANDREWS, K. Orangutan pantomime: Elaborating the message. **Biology Letters**, v. 7, n. 4, p. 627-630, 2011.
- RUSSON, A.; ANDREWS, K.; B, H. Innovation and the Grain Problem. **Behavioral and Brain Sciences**, v. 30, n. 4, p. 422, 2007.
- SAKURA, O. Similarities and Varieties: A Brief Sketch on The Reception of Darwinism and Sociobiology in Japan. **Biology and Philosophy**, v. 13, p. 341-357, 1998.
- SAKURA, O.; MATSUZAWA, T. Flexibility of wild chimpanzee nut-cracking behavior using stone hammers and anvils: an experimental analysis. **Ethology**, v. 87, p. 237-248, 1991.
- SANTOS, L.; NISSEN, A. Rhesus Monkeys, Macaca Mulatta, know what Others Can and Cannot Hear. **Animal Behaviour**, v. 71, p. 1175-1181, 2006.
- SAPOLSKY, R. The Influence of Social Hierarchy on Primate Health. **Science**, v. 308, p. 648-652, 2005.
- SAVAGE-RUMBAUGH, E. **Ape Language: From Conditioned Response to Symbol**. New York: Columbia University Press, 1986.
- SAVAGE-RUMBAUGH, E.; RUMBAUGH, D. Perspectives on Consciousness, Language, and Other Emergent Processes in Apes and Humans. *In*: HAMEROFF, S.; KASZNIAK, A.; A. (ed.). **Toward a Science of Consciousness II**. Scott: MIT Press, 1998. p. 533-549.
- SAVAGE-RUMBAUGH, E.; RUMBAUGH, D.; BOYSEN, S. Do Apes Use Language? **American Scientist**, v. 68, p. 49-61, 1980.
- SAVAGE-RUMBAUGH, E. *et al.* Reference: The Linguistic Essential. **Science**, v. 210, n. 4472, p. 922-925, 1980.
- SAYIGH, L. *et al.* Individual recognition in wild bottlenose dolphins: A field test using playback experiments. **Animal Behavior**, v. 57, p. 41-50, 1998.
- SCHAIK, C. *et al.* Orangutan Cultures and the Evolution of Material Culture. **Science**, v. 299, p. 102-105, 2003.
- SCHAIK, C.; NOORDWIJK, M.; WICH, S. Innovation in wild Bornean orangutans (*Pongo pygmaeus wurmbii*). **Behaviour**, v. 143, n. 7, p. 839-876, 2006.

- SCHUSTERMAN, R.; KRIEGER, K. California Sea Lions are Capable of Semantic Comprehension. **The Psychological Record**, v. 34, p. 3-23, 1984.
- SEARLE, J. Animal Minds. **Midwest Studies in Philosophy**, v. 19, n. 1, p. 206-219, 1994.
- SEYFARTH, R.; D. L., C.; MARLER, P. Vervet monkey alarm calls: Semantic communication in a free-ranging primate. **Animal Behavior**, v. 28, p. 1070-1094, 1980.
- SHANKER, S.; KING, B. The Emergence of a New Paradigm in Ape Language Research. **Behavioral and Brain Sciences**, v. 25, n. 5, p. 605-620, 2002.
- SHEETS-JOHNSTONE, M. Taking Evolution Seriously. **American Philosophical Quarterly**, v. 29, p. 343-352, 1992.
- SHETTLEWORTH, S. **Cognition, Evolution, and Behavior**. New York: Oxford University Press, 2010a.
- SHETTLEWORTH, S. Clever animals and killjoy explanations in comparative psychology. **Trends in Cognitive Science**, v. 14, n. 11, p. 477-481, 2010b.
- SHETTLEWORTH, S.; SUTTON, J. Do Animals Know What They Know? *In*: HURLEY, S.; NUDDS, M. **Rational Animals?** Oxford: Oxford University Press, 2006. p. 235-246
- SHUMAKER, R. **Observational Learning in Orangutans**. George Mason University, 1997.
- SHUMAKER, R.; SWARTZ, K. When Traditional Methodologies Fail: Cognitive Studies of Great Apes. *In*: BEKOFF, M.; ALLEN, C.; BURDHARDT, G. (ed.). **The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition**. Cambridge, MA: MIT Press, 2002. p. 335-345.
- SILK, J. *et al.* Chimpanzees Are Indifferent to The Welfare Of Unrelated Group Members. **Nature**, v. 437, p. 1357-1359, 2005.
- SKIDMORE, J. Duties to Animals: The Failure of Kant's Moral Theory. **The Journal of Value Inquiry**, v. 35, p. 541-559, 2001.
- SKINNER, B. The Generic Nature of the Concepts of Stimulus and Response. **The Journal of General Psychology**, v. 12, p. 40-65, 1935.
- SLOCOMBE, K. Food-Associated Calls in Chimpanzees: Responses to Food Types or Food Preferences? **Animal Behaviour**, v. 72, p. 989-999, 2006.
- SLOCOMBE, K.; ZUBERBÜHLER, K. Functionally Referential Communication in a Chimpanzee. **Current Biology**, v. 15, p. 1779-1784, 2005.
- SMET, A.; BYRNE, R. W. African Elephants Can Use Human Pointing Cues to Find Hidden Food. **Current Biology**, v. 23, n. 20, p. 2033-2037, 2013.

- SMITH, J.; SCHULL, J. The Uncertain Response in the Bottlenose Dolphin. **Journal of Experimental Psychology General**, v. 124, p. 391-408, 1995.
- SMITH, P. On Animal Beliefs. **Southern Journal of Philosophy**, v. 20, p. 503-512, 1982.
- SOBER, E. Comparative Psychology meets Evolutionary Biology: Morgan's Canon and Cladistic Parsimony. *In*: DATSON, L.; MITMAN, G. (ed.). **Thinking with Animals: New Perspectives on Anthropomorphism**. New York: Columbia University Press, 2005. p. 85-99.
- SOBER, E. Anthropomorphism, Parsimony, and Common Ancestry. **Mind & Language**, v. 27, n. 3, p. 229-238, 2012.
- SOBER, E. **Ockham's Razors: A User's Manual**. Cambridge: Cambridge University Press, 2015.
- SOLE, L.; SHETLEWORTH, S.; BENNETT, P. Uncertainty in Pigeons. **Psychonomic Bulletin and Review**, v. 10, p. 738-745, 2003.
- SPELKE, E. Physical Knowledge in Infancy: Reflections of Piaget's Theory. *In*: CAREY, S.; GELMAN, R. (ed.). **The Epigenesis of Mind: Essays on Biology and Cognition**. Hillsdale, NJ: Erlbaum, 1991. p. 37-61.
- SPERBER, D.; WILSON, D. **Relevance: Communication and Cognition**. Harvard University Press, 1986.
- STEGMANN, U. **Animal Communication Theory: Information and Influence**. Cambridge: Cambridge University Press, 2013.
- STICH, S. Do Animals Have Beliefs? **Australasian Journal of Philosophy**, v. 57, p. 15-28, 1978.
- STRUHSAKER, T. Auditory Communication Among Vervet Monkeys. *In*: ALTMANN, S. (ed.). **Social Communication among Primates**. Chicago: University of Chicago Press, 1967. p. 281-324.
- SUCHAK, M. *et al.* Ape duos and trios: spontaneous cooperation with free partner choice in chimpanzees. **PeerJ**, v. 2, p. 417, 2014.
- SWARTZ, K.; EVANS, S. Not All Chimpanzees Show Self-Recognition. **Primates**, v. 32, p. 483-496, 1991.
- SWARTZ, K.; SARAUW, D.; EVANS, S. Comparative Aspects of Mirror Self-Recognition in Great Apes. *In*: PARKER, T.; MITCHELL, R.; MILES, H. (ed.). **The Mentalities of Gorillas and Orangutans in Comparative Perspective**, 1999. p. 283-294.
- TERRACE, H. *et al.* Can an Ape Create a Sentence? **Science**, v. 206, n. 4421, p. 891-902, 1979.



- TETZLAFF, M.; REY, G. Systematicity and intentional realism in honeybee navigation".  
*In: LURZ, R. W. (ed.). **The Philosophy of Animal Minds***. Cambridge: Cambridge University Press, 2009. p. 72-88.
- THOMPSON, V.; PROWSE TURNER, J.; PENNYCOOK, G. Intuition, reason, and metacognition. **Cognitive Psychology**, v. 63, n. 3, p. 107-140, 2011.
- THORNDIKE, E. **Animal Intelligence**. New York: Macmillan, 1911.
- TINBERGEN, N. On Aims and Methods of Ethology. **Zeitschrift für Tierpsychologie**, v. 20, p. 410-433, 1963.
- TOMASELLO, M. Do Apes Ape? *In: HEYES, C.; GALEF, B. (ed.). **Do Apes Ape?*** London: Academic, 1996. p. 319-346.
- TOMASELLO, M. Emulation Learning and Cultural Learning. **Behavioral and Brain Sciences**, v. 21, p. 703-704, 1998.
- TOMASELLO, M. **The Cultural Origins of Human Cognition**. Cambridge, MA: Harvard University Press, 1999.
- TOMASELLO, M. Chimpanzees Versus Humans: It's Not That Simple. **Trends-in-Cognitive-Sciences**, v. 7, p. 239-240, 2003.
- TOMASELLO, M. **Origins of Human Communication**. Cambridge, Mass: MIT Press, 2008.
- TOMASELLO, M. **A Natural History of Human Morality**. Cambridge, MA: Harvard University Press, 2016.
- TOMASELLO, M.; CALL, J.; HARE, B. Chimpanzees Understand Psychological States-The Question is Which Ones and to What Extent. **Trends in Cognitive Science**, v. 7, p. 153-156, 2003.
- TOMASELLO, M.; CARPENTER, M. The Emergence of Social Cognition in Three Young Chimpanzees. **Monographs of the Society for Research in Child development**, v. 70, p. 1-131, 2005.
- TOMASELLO, M.; ZUBERBÜHLER, K. Primate Vocal and Gestural Communication. *In: BEKOFF, M.; ALLEN, C.; BURGHARDT, G. (ed.). **The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition***. Cambridge, MA: MIT Press, 2002.
- TSCHUDIN, A. *et al.* Comprehension of Signs by Dolphins (*Tursiops Truncatus*). **Journal of Comparative Psychology**, v. 115, n. 1, p. 100-105, 2001.
- ULLER, C. **Origins of Numerical Concepts: A Comparative Study of Human Infants and Nonhuman Primates**. Cambridge, MA: MIT Press, 1997.
- ULLER, C. Disposition to Recognize Goals in Infant Chimpanzees. **Animal Cognition**, v. 7, p. 154-161, 2004.

- VAN HOOFF, J. The Facial Displays of the Catarrhine Monkeys and Apes. In: MORRIS, D. (ed.). **Primate Ethology**. Chicago: Aldine, 1967. p. 7-78.
- VAN HOOFF, J. Comparative Approach to the Phylogeny of Laughter and Smile. In: HINDE, R. (ed.). **Non-Verbal Communication**. Cambridge University Press, 1972. p. 209-241.
- VAN HOOFF, J. A Structural Analysis of the Social Behaviour of a Semi-Captive Group of Chimpanzees. In: CRANACH, M.; VINE, I. (ed.). **Social Communication and Movement. European Monographs in Social Psychology 4**. London: Academic, 1973. p. 75-162.
- VARNER, G. **Personhood, Ethics, and Animal Cognition: Situating Animals in Hare's Two Level Utilitarianism**. New York: Oxford University Press, 2012.
- VOLTAIRE, F. **Philosophical Dictionary**. Tradução: BESTERMAN, T. New York: Penguin Classics, 1984.
- de WAAL, F. **Chimpanzee Politics: Power and Sex Among Apes**. London: Jonathan Cape, 1982.
- de WAAL, F. **Good Natured: The Origins of Right and Wrong in Humans and Other Animals**. Cambridge, MA: Harvard University Press, 1996.
- de WAAL, F. Anthropomorphism and Anthropodenial: Consistency in Our Thinking About Human and Other Animals. **Philosophical Topics**, v. 27, p. 225-280, 1999.
- de WAAL, F. Silent Invasion: Imanishi's Primatology and Cultural Bias in Science. **Animal Cognition**, v. 6, p. 293-299, 2003.
- de WAAL, F. **Primates and Philosophers: How Morality Evolved**. Princeton: Princeton University Press, 2006.
- de WAAL, F. **The Age of Empathy: Nature's Lessons for a Kinder Society**. Toronto: McClelland & Stewart, 2009.
- de WAAL, F.; BROSANAN, S. Monkeys reject unequal pay. **Nature**, v. 425, p. 297, 2003.
- de WAAL, F.; F. F. P. Towards a bottom-up perspective on animal and human cognition. **Trends in Cognitive Science**, v. 14, n. 5, p. 201-207, 2010.
- WARNEKEN, F. et al. Spontaneous Altruism by Chimpanzees and Young Children. **Public Library of Science Biology**, v. 5, n. 7, p. 1-7, 2007.
- WARNEKEN, F.; TOMASELLO, M. Altruistic Helping in Infants and Young Chimpanzees. **Science**, v. 311, n. 5765, 2006.
- WHITEN, A. When does smart behavior-reading become mind-reading? In: CARRUTHERS, P.; SMITH, P. (ed.). **Theories of Theories of Mind**. Cambridge: Cambridge University Press, 1995. p. -92.

- WHITEN, A.; BYRNE, R. Tactical deception in primates. **Behavioral and Brain Sciences**, v. 11, n. 2, p. 233-273, 1988.
- WHITEN, A. et al. How do Apes Ape? **Learning and Behavior**, v. 32, p. 36-52, 2004.
- WILSON, E. **Sociobiology**. Cambridge, MA: Belknap Press, 1975.
- WILSON, M. How did we get from there to here? An evolutionary perspective on embodied cognition. *In*: CALVO, P.; GOMILA, T. (ed.). **Handbook of Cognitive Science: An Embodied Approach**. San Diego: Elsevier, 2008.
- WIMMER, H.; PERNER, J. Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. **Cognition**, v. 13, p. 103-128, 1983.
- WIT, D.; DICKINSON, A. Associative theories of goal-directed behavior: A case for animal-human translational models. **Psychological Research**, v. 73, n. 4, p. 463-477, 2009.
- WITTIG, R. *et al.* Kin-mediated reconciliation substitutes for direct reconciliation in female baboons. **Proceedings of the Royal Society B: Biological Sciences**, v. 274, p. 1109-1115, 2007.
- WYNNE, C. The perils of anthropomorphism. **Nature**, v. 428, n. 606, p. 606, 2004.
- WYNNE, C. What are animals? Why anthropomorphism is still not a scientific approach to behavior. **Comparative Cognition and Behavior Reviews**, v. 2, p. 125-135, 2007.
- WYNNE, C.; UDELL, M. **Animal Cognition: Evolution, Behavior and Cognition**. Basingstoke, Hampshire; New York: Palgrave Macmillan, 2013.
- YAMAMOTO, S.; TANAKA, M. How did altruism and reciprocity evolve in humans? Perspectives from experiments on chimpanzees. **Interaction Studies**, v. 10, n. 3, p. 150-182, 2009.
- ZUBERBÜHLER, K. Referential Labeling in Diana Monkeys. **Animal Behavior**, v. 59, p. 917-927, 2000.
- ZUBERBÜHLER, K. Predator-Specific Alarm Calls in Campbell's Guenons. **Behavioral Ecology and Sociobiology**, v. 50, p. 414-422, 2001.
- ZUBERBÜHLER, K.; D.L., C.; M, S. R. Conceptual semantics in a nonhuman primate. **Journal of Comparative Psychology**, v. 113, p. 33-42, 1999.
- ZUBERBÜHLER, K.; R., N.; M, S. R. Diana monkeys long distance calls: Messages for conspecifics and predators. **Animal Behavior**, v. 53, p. 589-604, 1997.

## Sobre o organizador e os tradutores

**Bruno Malavolta e Silva** (tradutor). Realizou doutorado (2021), mestrado (2016) e bacharelado (2014) em filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS). Atua nas áreas de filosofia da ciência e epistemologia, com enfoque no debate sobre realismo científico.

**Eros Moreira de Carvalho** (organizador e tradutor). É Professor Associado do Departamento de Filosofia da Universidade Federal do Rio Grande do Sul. É Bolsista de Produtividade do CNPq. Graduiu-se em Filosofia pela UFMG e em Ciência da Computação pela UFPR. Obteve o título de mestre e o de doutor pela Universidade Federal de Minas Gerais (UFMG). É membro dos grupos de pesquisa *Enactive Cognition & Narrative Practices Research Group* (University of Wollongong), *Cognição, Linguagem, Enativismo e Afetividade* (CLEA) e coordena o *Grupo de Estudos e Pesquisa em Enativismo e Psicologia Ecológica* (GEPEPE). Suas áreas de interesse abrangem a filosofia das ciências cognitivas, a filosofia da percepção e a epistemologia. Sua pesquisa tem abordado temas da epistemologia e da filosofia da percepção, com base na concepção ecológica e enativa da mente.

**Jeferson Diello Huffermann** (tradutor). Doutorando do Programa de Pós-graduação em Filosofia da Universidade Federal do Rio Grande do Sul (UFRGS). Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Possui licenciatura e mestrado em Filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS). Áreas de atuação: Filosofia da Ciências Cognitivas, Filosofia da Ciência, Filosofia da Mente, Epistemologia.

**Jessica Nunes Vergara** (tradutora). Licenciada (2011), mestra (2019) e doutoranda em Filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS). Sua dissertação de mestrado, resultante da pesquisa orientada pelo Prof. Dr. Eros Moreira de Carvalho e financiada pelo CNPq, recebeu o Prêmio ANPOF 2020

promovido pela Associação Nacional de Pós-Graduação em Filosofia. É Técnica em Assuntos Educacionais no campus Charqueadas, do Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSUL). Membro do Grupo de Estudos e Pesquisa em Enativismo e Psicologia Ecológica (GEPEPE/UFRGS). Pesquisa, especialmente, com base na abordagem enativista, temas em Filosofia da Mente, Filosofia da Percepção e Metafísica.

**Pedro Maggi Rech Noguez** (tradutor). Bacharel (2018) e Mestre (2021) em Filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS). É pesquisador de doutorado na mesma instituição. Tem interesse em Filosofia da Linguagem, Filosofia da Ciência, Wittgenstein e Filosofia das Ciências Cognitivas.

**Rodrigo de Ulhôa Canto Reis** (tradutor). Possui graduação em Filosofia (2012) pela Universidade Federal do Rio Grande do Sul (UFRGS), mestrado (2015) e doutorado em Filosofia (2021) pela mesma universidade. Atua, principalmente, nos seguintes temas: História da Filosofia Contemporânea, Metafísica, Filosofia da Mente, Filosofia da Linguagem e Filosofia das Ciências Cognitivas.

**Thiago Gruner** (tradutor). Mestrando em Filosofia pela Universidade Federal do Rio Grande do Sul (UFRGS) e licenciado em Filosofia pela mesma universidade. É professor do Colégio Anchieta de Porto Alegre, da Rede Jesuíta de Educação. Áreas de atuação: Filosofia da Cognição, Filosofia da Arte e Ensino de Filosofia.

## Sobre os revisores

**César Fernando Meurer**. Possui graduação em filosofia (La Salle, 2006), mestrado em educação nas ciências (Unijuí, 2009), mestrado em filosofia (Unisinos, 2012) e doutorado em filosofia (Unisinos, 2016). Fez estágio pós-doutoral no departamento de Filosofia da Università Degli Studi di Milano (Milão, Itália) e no Instituto de Filosofia da Universidade Federal de Uberlândia (Uberlândia, Minas

Gerais). Atualmente, é professor no Centro de Ciências Naturais e Humanas da UFABC, com atuação no bacharelado em filosofia e nos bacharelados interdisciplinares. Suas pesquisas concentram-se em questões filosóficas, científicas e educacionais relativas à mente, à linguagem e ao tempo.

**Filipe Lazzeri.** É Professor Adjunto na Universidade Federal de Goiás (UFG). Obteve doutorado em Filosofia na Universidade de São Paulo (USP), com estágio de pesquisa na University of Miami. Possui mestrado e graduação em Filosofia pela Universidade de Brasília (UnB). Entre 2021 e 2022, realiza visita acadêmica (pós-doutorado) junto ao Programa de Pós-Graduação em Psicologia, da Universidade Federal do Espírito Santo (UFES).

**Nara M. Figueiredo.** É pesquisadora do Centro de Lógica, Epistemologia e História da Ciência da Universidade Estadual de Campinas (CLE-Unicamp). Bacharel e licenciada em filosofia pela mesma universidade, mestre e doutora em filosofia pela Universidade de São Paulo (USP), possui experiência internacional de pesquisa doutoral e pós-doutoral. Preza pela pesquisa interdisciplinar e pela atuação das mulheres na filosofia e nas ciências. Suas áreas de atuação são filosofia da mente, filosofia da linguagem e epistemologia. Sua pesquisa atual é em filosofia da cognição, com foco em enativismo e linguagem.

**Sofia Stein.** No âmbito do doutorado (USP), desenvolveu pesquisa na Universität Bielefeld (1998), Alemanha, acerca das obras de Rudolf Carnap e Willard Quine e, na University of Pittsburgh (2011), investigou a semântica naturalizada, sob a supervisão de John McDowell. Publicou *Van Orman Quine: epistemologia, semântica e ontologia* (2009) e traduziu o clássico quineano *Word and Object* (1960) (*Palavra e Objeto*, ed. Vozes, 2010). É uma das fundadoras do Laboratório de Filosofia Experimental e Estudos da Cognição (Labfic), da Unisinos (2012), primeiro laboratório de filosofia experimental do Brasil. Foi presidente da Sociedade Brasileira de Filosofia Analítica (SBFA, 2019-2020). Atualmente, é bolsista de produtividade em pesquisa 1D, do CNPq, e professora adjunta do Programa de Pós-graduação em Filosofia da Unisinos (2009-).





DISSERTATIO  
FILOSOFIA