Editora UFPel

Série Investigação Filosófica

Textos selecionados de

Filosofia da Psicologia Moral

Rodrigo Cid
Sagid Salles
(Organizadores)

DISSERTATIO
FILOSOFIA

Textos Selecionados de Filosofia da Psicologia Moral

Série Investigação Filosófica

TEXTOS SELECIONADOS DE FILOSOFIA DA PSICOLOGIA MORAL

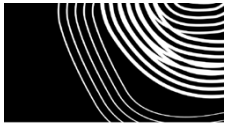
Rodrigo Cid
Sagid Salles
(Organizadores)

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 Agronômicas: Victor Fernando Büttow Roll
Representantes da Área das Ciências Exatas e da Terra: Eder João Lenardão
Representantes da Área das Ciências Biológicas: Rosângela Ferreira Rodrigues
Representante da Área das Engenharias e Computação: Reginaldo da Nóbrega Tavares
Representantes da Área das Ciências da Saúde: Fernanda Capella Rugno
Representante da Área das Ciências Sociais Aplicadas: Daniel Lena Marchiori Neto
Representante da Área das Ciências Humanas: Charles Pereira Pennaforte
Representantes da Área das Linguagens e Artes: Lúcia Bergamaschi Costa Weymar

EDITORA DA UFPEL
Chefia: Ana da Rosa Bandeira (Editora-chefe)
Seção de Pré-produção: Isabel Cochrane (Administrativo)
Seção de Produção: Suelen Aires Böettge (Administrativo)
Anelise Heidrich (Revisão)
Ingrid Fabiola Gonçalves (Diagramação)
Seção de Pós-produção: Madelon Schimmelpfennig Lopes (Administrativo)
Morgana Riva (Assessoria)



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

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

DIREÇÃO DO IFISP
Prof. Dr. João Hobuss

CHEFE DO DEPARTAMENTO DE FILOSOFIA
Prof. Dr. Juliano Santos do Carmo

Série Investigação Filosófica

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

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

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

ORGANIZADORES DO VOLUME
Rodrigo Cid (UNIFAP)
Sagid Salles (UESC)

TRADUTORES E REVISORES
César A. M. Alencar (UNIFAP)
Rafael Martins (UNIEDUK)
Sérgio Farias de Souza Filho (UFRJ)
Sagid Salles (UESC)
Aluízio Couto (UFMG)
Vinícius Brito Barros (UNIFAP)
Rafael Cavalcanti de Souza (UNICAMP)
L.H. Marques Segundo (UFSC)
Roseline Crippa (UNAR)

CRÉDITO DA IMAGEM DE CAPA
BOSCH, H. O Jardim das Delícias Terrenas, 1510-1515. Museu do Prado, Madri, ES (Domínio Público).

GRUPO DE PESQUISA INVESTIGAÇÃO FILOSÓFICA (UNIFAP/CNPq)

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

MEMBROS PERMANENTES DO GRUPO

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

© Série Investigação Filosófica, 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 psicologia moral.
[recurso eletrônico] Organizadores: Rodrigo Cid, Sagid Salles – Pelotas: NEPFIL Online, 2022.

366p. - (Série Investigação Filosófica).

Modo de acesso: Internet
<wp.ufpel.edu.br/nepfil>
ISBN: 978-65-86440-19-5

1. Filosofia. 2. Psicologia Moral I. Cid, Rodrigo. II. Salles, Sagid.

COD 100

Para maiores informações, visite o site wp.ufpel.edu.br/nepfil

SUMÁRIO

Sobre a série Investigação Filosófica **14**
Introdução **15**

(I) Psicologia Moral: Aproximações Empíricas **16**
1. Introdução: O que é psicologia moral? 17
2. Experimentos de pensamento e os métodos da ética 20
3. Responsabilidade Moral 26
4. Ética da virtude e ceticismo sobre o caráter 34
5. Egoísmo x Altruísmo 43
6. Discordância moral 47
7. Conclusão 60
Referência bibliográfica 60

(II) Raciocínio Moral **76**
1. A Importância Filosófica do Raciocínio Moral 77
1.1. Definindo "Raciocínio Moral" 77
1.2. Desafios Empíricos ao Raciocínio Moral 79
1.3. Situando o Raciocínio Moral 80
1.4. Obtendo Conhecimento Moral a partir do Estudo do Raciocínio Moral 83
1.5. Quão distinto é o Raciocínio Moral do Raciocínio Prático em Geral? 84
2. Questões Filosóficas Gerais Sobre Raciocínio Moral 86
2.1. Abordagem Moral 86
2.2. Princípios Morais 88
2.3. Selecionando Quais Considerações São Mais Relevantes 91
2.4. Raciocínio Moral e Psicologia Moral 95
2.5. Modelagem de Considerações Morais Conflitantes 100
2.6. Aprendizagem Moral e a Revisão das Teorias Morais 107
2.7. Como Podemos Raciocinar Moralmente Uns Com Os Outros? 109

Referência bibliográfica 111

(III) Motivação Moral **117**
1. O Fenômeno Básico da Motivação Moral 119
2. Motivação Moral e a Natureza das Propriedades Morais 121
3. Juízo Moral e Motivação 126
3.1. Humianismo *vs.* Anti-Humianismo 126
3.2. Internismo *vs.* Externismo 132
4. Motivação Moral e Metaética 142
5. Motivação Moral e Psicologia Experimental 145
Referência bibliográfica 151

(IV) Sentimentalismo Moral **157**
1. Os muitos sentimentalismos morais 158
2. Sentimentalismo explicativo 160
2.1. Teorias do senso moral 160
2.2. Teorias baseadas na simpatia/empatia 165
3. Sentimentalismo de Juízo 170
3.1. Sentimentalismo não cognitivista 170
3.2. Expressivismo 172
3.3. Sentimentalismo Cognitivo: Subjetivismo e Disposicionalismo Ideal 174
3.4. Sentimentalismo cognitivista: a teoria da sensibilidade e o neossentimentalismo 179
3.5. Abordagens Híbridas e Pluralistas 180
4. Metafísica Moral 183
4.1. Subjetivismo e Relativismo 184
4.2. Disposicionalismo Ideal 185
4.3. A Teoria da Sensibilidade e o Neossentimentalismo 189
5. Sentimentalismo Epistêmico 192
5.1. O Modelo Perceptual 194
5.2. Teorias do Rastreamento 202
5.3. Entendimento Moral e Testemunho 203
6. Uma Variedade de Sentimentalismos Morais (Reprise) 206
Referência bibliográfica 207

(V) Psicologia Evolutiva **216**
1. Psicologia Evolutiva: Uma tradição de pesquisa entre as várias abordagens biológicas para explicar o comportamento humano 217
2. Teoria e métodos da Psicologia Evolutiva 219
3. A hipótese da modularidade massiva 222
4. Filosofia da Biologia vs. Psicologia Evolutiva 228
5. Psicologia Moral e Psicologia Evolutiva 235
6. Natureza Humana 236
7. Aplicações da Psicologia Evolutiva e Prospectos para o Debate Futuro 238
Referência bibliográfica 241

(VI) A Psicologia da Cognição Normativa **250**
1. Uma Capacidade Psicológica Dedicada a Normas 252
1.1. *Background*: Evolução e Considerações Distais 257
1.2. Psicologia e Explicações Proximais: Ferramentas Teóricas e Dimensões 261
2. Pesquisa Empírica 270
2.1. Sociologia, Antropologia e Psicologia Cultural 271
2.2. Economia Comportamental 273
2.3. Psicologia Desenvolvimentista e Comparativa 275
3. Cognição Normativa e Moralidade 279
Suplemento: descrição estendida da figura 282
Referência bibliográfica 282

(VII) Filosofia Moral Experimental **303**
1. Introdução e História 304
1.1. O Que São Experimentos? 306
1.2. Quais Tipos de Questões e Dados? 308
2. Intuições Morais e Análise Conceitual 310
2.1. O Programa Negativo na Filosofia Moral Experimental 311
2.2. O Programa Positivo na Filosofia Moral Experimental 312
2.3. Um Exemplo: Intencionalidade e Responsabilidade 314
2.4. Outro Exemplo: A Analogia Linguística 319
3. Caráter, Bem-estar e Emoção 325
3.1. Caráter e Virtude 325
3.2. Bem-estar, Felicidade e a Vida Boa 330
3.3 Emoção 334

4. Metaética e Filosofia Moral Experimental 337
4.1. Metaética de Senso Comum e Realismo Moral 338
4.2. Desacordo Moral 340
4.3. Linguagem Moral 341
5. Críticas à Filosofia Moral Experimental 343
5.1. Problemas com o *Design* e Interpretação Experimental 343
5.2. Problemas Filosóficos 347
Referência bibliográfica 348

Sobre organizadores e tradutores 363

SOBRE A SÉRIE INVESTIGAÇÃO FILOSÓFICA

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

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

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

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

INTRODUÇÃO

A compreensão adequada de uma área de investigação tipicamente envolve pelo menos três elementos: problemas, teorias e argumentos. Aqueles que desejam compreender a área conhecida como "psicologia moral" precisam ter uma boa ideia dos problemas que orientam os filósofos e cientistas engajados nela, das teorias que são propostas para resolver estes problemas e dos argumentos a favor e contra estas teorias. Acreditamos que a presente seleção de textos fornecerá exatamente isto. Não levará muito tempo para que o leitor perceba que a psicologia moral é um campo interdisciplinar, que envolve, principalmente, a filosofia (incluindo a recente filosofia experimental) e a psicologia. Trata-se de um belíssimo exemplo de como duas disciplinas distintas podem se informar mutuamente, de modo a progredir juntas rumo à solução de problemas espinhosos. É, também, um exemplo de como a investigação conceitual pode andar lado a lado com a investigação empírica, a despeito de eventuais conflitos entre elas. O objeto da psicologia moral é o comportamento humano em contextos morais, um tema relativamente novo, bastante complexo, e acerca do qual ainda há pouca coisa em português. Esperamos que este volume seja um bom começo para os curiosos, e que sirva de inspiração para trabalhos futuros. Uma boa leitura a todos!

Sagid Salles
Rodrigo Cid
Organizadores do Livro

(I) Psicologia Moral: Aproximações Empíricas*

Autoria: John Doris, Stephen Stich, Jonathan Phillips e Lachlan Walmsley
Tradução: Cesar A. M. de Alencar
Revisão: Vinícius Brito Barros

A psicologia moral investiga a performance humana em contextos morais e indaga como esses resultados podem impactar o debate em teoria ética. Tal investigação é necessariamente interdisciplinar, valendo-se ao mesmo tempo dos recursos empíricos das ciências humanas e dos recursos conceituais da ética filosófica. O presente artigo discute vários tópicos que ilustram esse tipo de investigação: experimentos mentais, responsabilidade, caráter, egoísmo x altruísmo e discordância moral.

---

* DORIS, J.; STICH, S.; PHILLIPS, J.; WALMSLEY, L. Moral Psychology: Empirical Approaches. In: ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2020. Disponível em: https://plato.stanford.edu/archives/spr2020/entries/moral-psych-emp/. Acesso em: 03 maio 2022.

The following is the translation of the entry on Moral Psychology: Empirical Approaches by Doris, John; Stephen Stich; Jonathan Phillips; e Lachlan Walmsley 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/moral-psych-emp/. 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 a https://plato.stanford.edu/entries/moral-psych-emp/. We'd like to thank the Editors of the Stanford Encyclopedia of Philosophy, mainly Prof. Dr. Edward Zalta, for granting permission to translate and to publish this entry.

1. Introdução: O que é psicologia moral?

A psicologia moral contemporânea – que se constitui pelo estudo do pensamento e comportamento humano em contextos éticos – é decididamente interdisciplinar: os psicólogos se baseiam livremente em teorias filosóficas a fim de auxiliarem a estruturação de sua pesquisa empírica, enquanto os filósofos se baseiam livremente em descobertas empíricas da psicologia a fim de auxiliarem a estruturação de suas teorias[1] .

Embora essa expressiva interdisciplinaridade seja um desenvolvimento bastante recente (a maioria dos trabalhos relevantes, com poucas exceções, foram publicados no último quarto de século), não deveríamos tomá-lo como surpreendente. Desde a antiguidade até o presente, os filósofos jamais se mostraram tímidos em fazer afirmações empíricas, e muitas dessas afirmações têm sido acerca da psicologia humana (Doris; Stich, 2005). Desse modo, não surpreende que, com o surgimento da psicologia científica no último século e meio, alguns desses filósofos se propusessem comparar seu trabalho com as descobertas sistemáticas de psicólogos (tomando, auspiciosamente, um cuidado especial para evitar serem enganados por controvérsias científicas; *vide* DORIS, 2015, cap. 3; Machery; Doris, no prelo).

De modo semelhante, ao menos desde o fim do behaviorismo, os psicólogos têm se mostrado bastante interessados nos fenômenos normativos em geral, e nos fenômenos éticos em particular. Destarte, é normal que alguns deles busquem enriquecer seu arcabouço teórico com os recursos conceituais de um campo tão intensamente direcionado aos fenômenos normativos como o da ética filosófica. Por essa razão, a área demarcada pela "psicologia moral" envolve, rotineiramente, uma mistura de investigações empíricas e normativas desenvolvidas por filósofos e psicólogos – assumindo, cada vez mais, a forma de esforços colaborativos, envolvendo profissionais de ambas as áreas.

Para os filósofos, o interesse especial por essa investigação interdisciplinar assenta-se nos modos pelos quais a psicologia moral pode auxiliar o julgamento

---

[1] Para perspectivas mais gerais sobre essa literatura, vide Tibério (2015) e Alfano (2016). Para coleções úteis de artigos, vide Sinnott-Armstrong (2008a, 2008b, 2008c, 2014), Doris et al. (2010), Decety e Wheatley (2015) e Voyer e Tarantola (no prelo).

sobre teorias éticas concorrentes. A plausibilidade de uma psicologia moral associada não constitui, é claro, a única dimensão a partir da qual uma teoria ética pode ser avaliada; igualmente importantes são as questões **normativas** relacionadas a quão eficaz uma teoria se mostra quando comparada a convicções importantes sobre coisas como justiça, equidade e boa vida. Essas questões foram, e continuarão a ser, de importância central para a ética filosófica. No entanto, supõe-se comumente que uma teoria ética, comprometida com uma concepção empobrecida ou imprecisa da psicologia moral, está em séria desvantagem competitiva. Como argumentou, de modo assertivo, Bernard Williams (1973, 1985; *vide* Flanagan, 1991), uma concepção ética que recomenda relacionamentos, compromissos ou projetos de vida que entram em conflito com tipos de afeições que seriam razoavelmente esperadas no enraizamento e na rotina da vida real dos seres humanos é – no melhor dos casos – uma concepção ética com uma reivindicação muito tênue ao nosso assentimento.

Isso posto, aqueles problemas de escolha em teoria ética, que fazem referência à psicologia moral, podem ser enquadrados por duas indagações relacionadas:

> I Que afirmações empíricas sobre a psicologia humana os defensores de perspectivas concorrentes em teoria ética afirmam ou pressupõem?
> II Quão empiricamente bem fundamentadas são essas afirmações?

A primeira questão é própria aos estudos filosóficos: quais os compromissos psicológicos das várias posições em ética filosófica? A segunda questão nos leva além dos corredores dos departamentos de filosofia e do tipo de perguntas feitas, e às vezes respondidas, pelas ciências humanas, incluindo psicologia, antropologia, sociologia, história, ciências cognitivas, linguística e neurociência. Por isso, a psicologia moral contemporânea pode ser dita **metodologicamente pluralista**: pretende responder às questões filosóficas, mas de um modo empiricamente responsável.

Contudo, será por vezes difícil dizer quais afirmações em ética filosófica requerem fundamentação empírica. Isso se dá, em parte, por não estar claro às vezes se, e em que medida, uma controvérsia conta como empiricamente avaliável. Considere as questões relacionadas ao "funcionamento normal" em cuidados de saúde mental: respostas a essas questões são estatísticas ou avaliativas (Boorse, 1975; Fulford, 1989; Murphy, 2006)? Por exemplo, a saúde mental "normal" é,

simplesmente, a condição psicológica da maioria das pessoas, ou é uma condição de boa saúde mental? No primeiro caso, a questão é, ao menos em princípio, empiricamente determinável. Neste último caso, as questões devem ser decididas, se puderem ser decididas, por argumentos sobre valor.

Além disso, filósofos nem sempre se mostram explícitos sobre se, e até que ponto, estão fazendo afirmações empíricas. Por exemplo, suas representações de um caráter moral têm como objetivo identificar características psicológicas de pessoas reais ou articular ideais que não precisam ser instanciados em psicologias humanas reais? Essas questões certamente serão complicadas, devido à inevitável diversidade de opiniões filosóficas.

Deste modo, em todo caso, a primeira tarefa será documentar cuidadosamente as afirmações empiricamente avaliáveis de uma teoria, sejam elas explícitas ou, como frequentemente é o caso, tácitas. Uma vez localizadas as afirmações aptas para uma avaliação empírica, a questão passa a ser identificar quaisquer literaturas empíricas relevantes. O próximo trabalho será avaliar essas literaturas, na tentativa de determinar quais conclusões podem ser, com responsabilidade, obtidas delas. A ciência, em particular a ciência social, sendo o que é, torna muitas conclusões provisórias; o psicólogo, enquanto filósofo moral, deve estar preparado para julgar controvérsias em outros campos ou oferecer conjecturas informadas sobre descobertas futuras. Frequentemente, o registro empírico estará inevitavelmente incompleto. Em casos como esses, os filósofos podem ser forçados a se envolverem em conjecturas disciplinadas empiricamente, ou mesmo a se envolverem no próprio trabalho empírico, como alguns filósofos estão começando a fazer[2] .

Quando as posições filosóficas forem isoladas e as literaturas empíricas supostamente relevantes forem avaliadas, será possível começarmos a avaliar a plausibilidade da psicologia moral filosófica: seria o quadro especulativo do funcionamento psicológico, relacionado a alguma região da teoria ética, compatível com o enquadramento empírico que emerge da observação sistemática? Em suma, o quadro filosófico é **empiricamente adequado**? Se ficar determinado que a concepção filosófica é empiricamente adequada, o resultado está **justificado**. Inversamente, se a psicologia moral filosófica em questão for considerada

---

[2] Uma visão geral sobre esse trabalho pode ser encontrada no verbete sobre **filosofia moral experimental**. Para alguns artigos recentes, vide Sarkissian e Wright (2014).

empiricamente **inadequada**, o resultado será uma alteração **revisional** e convincente, ou mesmo uma rejeição, daqueles elementos da teoria filosófica que pressupõem a psicologia moral problemática. O processo geralmente será **comparativo**. A escolha da teoria em psicologia moral, como qualquer escolha de teoria, envolve trocas e, embora uma abordagem empiricamente mal fundamentada possa não ser eliminada de forma decisiva da disputa apenas em bases empíricas, pode vir a ser vista como menos atraente do que opções teóricas com fundamentos empíricos mais firmes.

Os ventos, que conduzem essa **polinização cruzada** que descrevemos entre as disciplinas, não sopram em uma única direção. Como filósofos que escrevem para uma enciclopédia de filosofia, estamos naturalmente preocupados com as maneiras pelas quais a pesquisa empírica pode moldar ou reformar a ética filosófica. Mas a reflexão filosófica pode, igualmente, influenciar a pesquisa empírica, uma vez que tal pesquisa é frequentemente conduzida por suposições filosóficas que podem ser mais ou menos filosoficamente seguras. Os melhores diálogos interdisciplinares devem, então, beneficiar ambas as partes. Para ilustrar o processo dialético que descrevemos, consideraremos uma variedade de tópicos da psicologia moral. Nossas principais preocupações serão filosóficas: quais são alguns dos problemas mais importantes da psicologia moral filosófica e como eles podem ser resolvidos? No entanto, como a natureza híbrida de nosso tópico nos convida a fazer, buscaremos essas questões em um espírito interdisciplinar, e temos esperança de que nossas observações também envolvam cientistas interessados. Esperamos que o resultado seja um amplo juízo acerca dos problemas e métodos que estruturarão a pesquisa em psicologia moral durante o século XXI.

2. Experimentos de pensamento e os métodos da ética

"Bombas de intuição" ou "experimentos de pensamento" têm sido itens bem utilizados na caixa de ferramentas do filósofo contemporâneo (Dennett 1984, p. 17-18; Stuart *et al*., 2018). Costumeiramente, um experimento mental apresenta um exemplo, quase sempre hipotético, com o intuito de provocar certa resposta filosoficamente eficaz. Se um experimento mental for bem-sucedido, pode-se concluir que teorias concorrentes devem dar conta da resposta resultante. Essas respostas seguem tendo uma **função de evidência** para a escolha da teoria filosófica; dito de outro modo, elas podem ser entendidas como **dados** que as teorias concorrentes

devem acomodar[3]. Se as respostas éticas de um certo público apropriado a um experimento mental entrarem em conflito com a resposta que uma teoria prescreve para o caso, a teoria sofreu um contraexemplo.

A questão sobre quais respostas "contam" filosoficamente (ou quem é o público "apropriado") foi respondida de modos variados. Para muitos filósofos, contudo, o público pretendido em experimentos mentais parece estar relacionado a certo tipo de "pessoa comum". (*vide* Jackson, 1998, p. 118, p. 129; Jackson; Pettit, 1995, p. 22-29; Lewis, 1989, p. 126-129). É claro que as pessoas relevantes deverão possuir os conhecimentos cognitivos necessários para compreender o caso em questão; crianças muito pequenas provavelmente não serão um público ideal para experimentos mentais. Assim, alguns filósofos podem insistir que as respostas relevantes serão os julgamentos ponderados de pessoas com o treinamento necessário para se atentarem "ao que está filosoficamente em jogo". Mas se as respostas pretendem ajudar a decidir entre teorias concorrentes, os respondentes devem ser mais ou menos teoricamente neutros, e esse tipo de neutralidade provavelmente estará viciado pela educação filosófica. Surge um dilema. De um lado, sujeitos filosoficamente ingênuos podem ser pensados carecendo da erudição necessária para compreenderem os desafios filosóficos. De outro lado, com o aumento da sofisticação filosófica vem, muito provavelmente, a parcialidade filosófica; um público é ingênuo e o outro preconceituoso[4].

Ainda que não importe, ao certo, como o público filosoficamente relevante é especificado, há questões empíricas que devem ser abordadas na determinação da potência filosófica de um experimento mental. Em particular, ao decidir que peso

---

[3] Essa analogia científica não é exclusiva da exposição que fazemos. Singer (1974, p. 517; p. 493) compreende o método de equilíbrio reflexivo de Rawls (1971) como "levando-nos a pensar em nossos juízos morais particulares como sendo os materiais contra os quais as teorias morais precisam ser testadas [...]". Como nota Singer (1974, p. 493, n.3), Rawls (1951), em artigo anterior, deixou explícita a analogia com a escolha da teoria científica. Embora seja amplamente aceita a afirmação que diz serem as intuições usadas como evidência em filosofia (Pust, 2000; Sosa, 2007), ela foi contestada recentemente, sobretudo por Cappelen (2012) e Deustch (2015). Para uma leitura útil acerca desse debate, confira Nado (2016).

[4] Para outras informações sobre o apelo às intuições em "especialistas" filosóficos, confira os trabalhos de Deutsch (2015, p. 134-135), Stich e Tobia (2016, p. 14-17; 2018).

filosófico dar a uma resposta, os filósofos precisam determinar suas **origens**. Que características do **exemplo** estão implicadas em um determinado julgamento – as pessoas estão reagindo à substância do caso ou ao estilo de exposição? Que características do **público** estão implicadas em sua reação – diferentes grupos demográficos respondem ao exemplo de maneira diferente? Existem fatores no ambiente que estão afetando os julgamentos intuitivos das pessoas? A ordem em que as pessoas consideram os exemplos afeta seus julgamentos? Tais questões levantam a seguinte preocupação: julgamentos sobre experimentos mentais que lidam com questões morais podem ser fortemente influenciados por características **eticamente irrelevantes** do exemplo, do público, do ambiente ou da ordem de apresentação. Uma característica é eticamente relevante por ser uma questão para discussão filosófica. No entanto, determinar o estatuto de um experimento de pensamento específico também requer investigação empírica de suas características causalmente relevantes. Vamos agora descrever alguns exemplos de tal investigação.

Como parte de sua famosa pesquisa sobre "heurísticas e vieses" subjacentes ao raciocínio humano, Tversky e Kahneman (1981) propuseram aos participantes o seguinte problema:

> Imagine que os EUA estejam se preparando para o surto de uma doença asiática incomum, com a expectativa de que ela mate 600 pessoas. Dois programas alternativos para combater a doença foram propostos. Suponha que a estimativa científica exata das consequências de cada programa proposto sejam as seguintes:
>
> - Se o Programa A for adotado, 200 pessoas serão salvas.
> - Se o Programa B for adotado, há 1/3 de probabilidade de que 600 pessoas sejam salvas e 2/3 de probabilidade de que nenhuma pessoa seja salva.
>
> Um segundo grupo de participantes recebeu um problema idêntico, exceto pelos programas terem sido descritos da seguinte forma:
>
> - Se o Programa C for adotado, 400 pessoas morrerão.

- Se o Programa D for adotado, há 1/3 de probabilidade de que ninguém morra e 2/3 de probabilidade de que 600 pessoas morram.

Na primeira versão do problema, a maioria dos participantes pensou que o Programa A deveria ser adotado. Mas, na segunda versão, a maioria escolheu o Programa D, apesar de o resultado descrito em A ser idêntico ao descrito em C. A desconcertante implicação desse estudo assinala que as respostas éticas podem ser fortemente influenciadas pela maneira como os casos são descritos ou enquadrados. Ao que parece, essa sensibilidade no enquadramento constitui uma influência eticamente irrelevante em respostas éticas. A não ser que essa possibilidade possa ser eliminada com segurança, deve-se hesitar ao confiar nas respostas a um experimento mental a fim de julgar controvérsias teóricas. Tais possibilidades só podem ser eliminadas por meio de um trabalho empírico sistemático[5] .

Embora uma porcentagem relativamente pequena daquele trabalho empírico sobre "heurística e vieses" aborde diretamente o raciocínio moral, diversos filósofos que abordaram a questão (Horowitz, 1998; Doris; Stich, 2005; Sinnott-Armstrong, 2005; Sunstein, 2005) concordam que fenômenos, tais como os efeitos de enquadramento, são passíveis de aparecerem sutilmente implicados em respostas a exemplos comprometidos eticamente, e argumentam que esse estado de coisas deve provocar os filósofos a verem o método experimental de pensamento com considerável preocupação.

Passemos agora aos efeitos de ordenação. Em um estudo pioneiro, Petrinovich e O'Neill (1996) descobriram que as intuições morais dos participantes variavam conforme a ordem em que os experimentos mentais eram apresentados. Achados semelhantes chegaram a ser relatados por Liao *et al*. (2012), Wiegman *et al*. (2012) e Schwitzgebel, Cushman (2011; 2015). Os estudos de Schwitzgebel e Cushman são particularmente notáveis, por se proporem a explorar se os efeitos de ordem em intuições morais eram menores ou inexistentes nos filósofos profissionais.

---

[5] Certos autores, como Baron (1994) e Sunstein (2005), argumentam que as influências distorcidas de "heurísticas e vieses" divulgadas na literatura psicológica recente sobre raciocínio, julgamento e decisão estão difundidas na reflexão ética cotidiana. Para outras leituras dessa literatura psicológica relevante, confira Nisbett e Ross (1980), Kahneman et al. (1982), Barão (2001), Gilovich et al. (2002) e Kahneman (2011).

Surpreendentemente, eles descobriram que filósofos profissionais estavam também sujeitos aos efeitos de ordenação, ainda que os experimentos mentais utilizados fossem bem conhecidos na área. Schwitzgebel e Cushman também relatam que, em alguns casos, as intuições desses filósofos mostravam efeitos de ordem substanciais, diferente das intuições dos não-filósofos.

Características do público também podem afetar o resultado de experimentos mentais. Haidt e seus associados (1993, p. 613) descreveram histórias sobre "violações, inocentes porém repugnantes, de fortes normas sociais" para homens e mulheres, em posição alta e baixa no Status Socioeconômico (SES), na Filadélfia (EUA), em Porto Alegre e em Recife (ambos no Brasil). Segue um exemplo:

> Um homem vai ao supermercado uma vez por semana e compra uma galinha morta. Antes de cozinhá-la, porém, ele tem relações sexuais com ela. Então, ele a cozinha e a come. (Haidt *et al*., 1993, p. 617)

Participantes com SES mais baixo tendiam a "moralizar" comportamentos inocentes e repugnantes similares ao da história da galinha. Esses sujeitos mostravam-se mais inclinados que seus colegas com alto NSE a dizer que o ator deveria ser "impedido ou punido", além de mais inclinados a negar que tais comportamentos estariam "certos", se fossem costumeiros em um determinado país (Haidt *et al*., 1993, p. 618-619). Não se trata de os participantes com SES mais baixo estarem equivocados na moralização que fazem de tais comportamentos, enquanto a civilidade dos colegas de SES mais alto signifique uma resposta mais racionalmente defensável. A dificuldade está em decidir qual das respostas conflitantes, caso haja, será adequada à função de restringir a teoria ética, porquanto ambas podem ser igualmente o resultado de fatores culturais mais ou menos arbitrários.

Geralmente, audiências filosóficas rejeitam moralizar os comportamentos repugnantes, e sem dúvidas compartilhamos dessa atitude tolerante. É claro, no entanto, que essas audiências – em virtude de aquisições educacionais, senão de suas carteiras de ações – possuem um SES decisivamente alto. O estudo de Haidt sugere que é um erro para um filósofo dizer, como faz Jackson (1998, p. 32, n. 4; p.37), que "minhas intuições revelam a concepção popular na medida em que tenho o razoável direito, como geralmente tenho, de me considerar alguém típico". A questão é: típico de qual demografia? As respostas éticas dos filósofos são determinadas pela substância filosófica dos exemplos, ou por idiossincrasias culturais

que são muito possivelmente consideradas eticamente irrelevantes? Vale dizer, até que tais possibilidades sejam descartadas por uma investigação empírica sistemática, o peso filosófico de um experimento mental estará aberto a questionamentos.

Recentemente, certo campo de pesquisas tem crescido ao noticiar que aqueles julgamentos evocados por experimentos de pensamento moral estão afetados por fatores circunstanciais que parecem ser completamente irrelevantes para o problema moral em questão. A presença de caixas de pizza sujas e um spray que solta cheiro de peido (Schnall *et al*., 2008a), o uso de sabão (Schnall *et al*., 2008b) ou um lenço antisséptico (Zhong *et al*., 2010), ou até mesmo a proximidade de um dispenser para higienizar as mãos (Helzer; Pizarro, 2011) foram relatados como influenciando as intuições morais. Tobia *et al*. (2013) descobriram que as intuições morais de estudantes e filósofos profissionais eram afetadas quando o questionário era pulverizado com spray desinfetante. Valdesolo e DeSteno (2006) noticiaram que assistir a um videoclipe humorístico pode ter um impacto substancial sobre as intuições morais dos participantes. E Strohminger *et al*. (2011) mostraram que ouvir diferentes tipos de clipes de áudio (comédia de *stand-up* ou histórias inspiradoras como as de um disco chamado *Chicken Soup for the Soul*) causa efeitos dissonantes para intuições morais.

De que maneira os teóricos morais deveriam reagir a descobertas como essas? Pode-se, é claro, evitar os experimentos mentais na teorização ética. Só que esse tipo de austeridade metodológica, embora não seja indesejável, traz um custo. Não obstante suas dificuldades, os experimentos mentais acabam sendo uma janela, em alguns casos a única janela acessível, para regiões importantes da experiência ética. Enquanto está desconectada dos pensamentos e sentimentos da vida ética cotidiana, a teoria ética corre o risco de ser "motivacionalmente inacessível", ou incapaz de engajar o interesse ético dos agentes que deveriam viver de acordo com os padrões normativos da teoria[6] . Felizmente, há uma outra possibilidade: continuar buscando o programa de pesquisa que investiga, sistematicamente, as respostas oferecidas às bombas de intuição. De fato, a ideia

---

[6] Rosati (1995) levanta preocupações análogas em relação a "relatos de informação completa", altamente idealizados, do que é bom: o agente ideal, totalmente informado, torna-se provavelmente tão diferente psicologicamente do agente real, informado de maneira escassa, que haveria pouca razão para pensar que o agente real seria motivado pelo que motiva o agente ideal.

tem sido submeter experimentos filosóficos de pensamento aos métodos críticos da psicologia social experimental. Se as investigações que empregam cenários experimentais diversos e populações de participantes revelam uma clara tendência nas respostas oferecidas, podemos começar a ter confiança de estarmos identificando certa convicção moral profunda e amplamente compartilhada. A discussão filosófica pode estabelecer quais convicções desse tipo devem servir como restrição à teoria moral, ao passo que respostas a experimentos de pensamento que a pesquisa empírica determina como carentes de tal solidez – tais como os suscetíveis à ordem, ao enquadramento ou aos efeitos circunstanciais, ou mesmo aqueles que admitem fortes influências de variações culturais – podem ser desconsiderados com segurança pelos teóricos éticos.

3. Responsabilidade moral

Um programa de pesquisa empírica filosoficamente instruído, similar àquele que descrevemos, é mais do que uma fantasia metodológica. Esse procedimento descreve com precisão um conjunto de programas de pesquisa disposto a influenciar os debates filosóficos através de pesquisas interdisciplinares.

Um dos exemplos mais antigos desse tipo de procedimento foi, em grande parte, inspirado pelos estudos de Knobe (2003a, 2003b, 2006), ao abordarem questões acerca da "moralidade popular" em função de tópicos abrangendo desde ação intencional até responsabilidade causal (*vide* Knobe, 2010). Esse trabalho inicial ajudou a estimular o desenvolvimento de um programa de pesquisa verdadeiramente interdisciplinar, com filósofos e psicólogos investigando a moralidade popular da vida cotidiana. Para tratamento mais completo desta pesquisa, *vide* verbete da SEF sobre **Filosofia Moral Experimental**.

Um outro debate filosófico, relacionado a esse, está interessado na compatibilidade do livre arbítrio e da responsabilidade moral com o determinismo. De um lado, incompatibilistas insistem que o determinismo (a visão de que os eventos são conjuntamente determinados por eventos antecedentes, governados por leis da natureza) é **incompatível** com a responsabilidade moral. Normalmente, tais argumentos chegam também a especificar qual capacidade singular é requerida para alguém ser responsável pelo seu próprio comportamento (por exemplo, que os agentes possuem possibilidades alternativas ao seu comportamento, ou que o agente é a fonte "última" de seu comportamento, ou ambos (Kane, 2002, p. 5; Haji,

2002, p. 202–203)[7] . De outro lado, compatibilistas argumentam que o determinismo e a responsabilidade moral são **compatíveis**, muitas vezes negando que a ação responsável requer do agente que ele tenha alternativas genuinamente abertas, ou rejeitando a condição última que exige o indeterminismo (ou as demandas impossíveis de autocriação). Em suma, compatibilistas sustentam que as pessoas podem ser responsabilizadas legitimamente mesmo havendo algum sentido em dizer que elas "não poderiam ter feito de outra forma", ou que elas não são a "razão final" de seu comportamento. Defensores dessas duas posições conflitantes permanecem relativamente entrincheirados, e alguns deles temem se tratar de um "impasse dialético" (Fischer, 1994, p. 83-85).

Nesse debate, um ponto crítico ocorre com a afirmação de que a posição incompatibilista apreende melhor os julgamentos morais populares acerca dos agentes cujas ações foram completamente determinadas (*vide* G. Strawson, 1986, p. 88; Smilansky, 2003, p. 259; Pereboom, 2001, p. XVI; O'Connor, 2000, p. 4; Nagel, 1986, p. 113, p. 125; Campbell, 1951, p. 451; Pink, 2004, p.12). É o caso do representante incompatibilista Robert Kane (1999, p. 218; 1996, p. 83-85), ao afirmar que, em sua experiência, "a maioria das pessoas comuns começa como incompatibilistas naturais", e, então, "acabam sendo afastadas desse incompatibilismo natural por hábeis argumentos de filósofos".

Como era de se esperar, alguns compatibilistas agiram rapidamente para afirmar o contrário. Por exemplo, Peter Strawson (1982) argumentou, de modo notável, que no âmbito das "relações interpessoais cotidianas" as pessoas não se veem assombradas pelo espectro do determinismo; tais preocupações metafísicas são irrelevantes para a experiência comum, além de expressarem "atitudes reativas" – raiva, ressentimento, gratidão, perdão e coisas semelhantes – associadas à avaliação da responsabilidade. Qualquer inquietação com o determinismo, insistiu Strawson, decorre de uma "metafísica apavorada" de filósofos, antes de ter algo a

---

[7] Há uma riqueza de variação terminológica e doutrinária aqui. Por exemplo, Haji usa "autenticidade" na circunstância onde Kane usa "razão final". Muitas vezes, tais questões são formuladas em termos de "liberdade" em vez de (ou junto com) "responsabilidade"; de fato, como observa Kane (2002, p. 4-5), quase sempre as questões não são distinguidas claramente. Concentramo-nos aqui inteiramente na atribuição de responsabilidade, e por isso nos omitimos sobre a relação (ou a falta de relação) entre essas questões e aquelas discutidas sob o título de "liberdade" ou "livre arbítrio".

ver com convicções incompatibilistas por parte das pessoas comuns. No entanto, os incompatibilistas costumam pensar que, historicamente, as intuições do cotidiano estão do seu lado; mesmo certos filósofos, com tendências compatibilistas, mostram-se dispostos a admitirem a posição incompatibilista ao se referirem a tendências de resposta "típicas" (*vide* VARGAS, 2005a, 2005b).

Nenhum dos lados, até onde ficamos sabendo, tem oferecido o bastante para uma evidência sistemática de padrões reais para juízos morais populares. Recentemente, contudo, um agora considerável programa de pesquisa começou a oferecer evidências empíricas acerca da relação entre determinismo e responsabilidade moral em juízos morais populares.

Inspirados pelo trabalho de Frankfurt (1988) e outros, Woolfolk, Doris e Darley (2006) levantaram a hipótese de que os observadores podem responsabilizar os agentes mesmo quando aqueles julguem que estes não poderiam ter agido de outra forma, desde que os agentes pareçam se "identificar" com seu comportamento. Grosso modo, a ideia seria que o agente se identifica com certo comportamento – sendo, portanto, responsável por ele – na medida em que ele o "abrace", quer dizer, realize-o "de todo o coração", não obstante serem possíveis alternativas genuínas para tal comportamento[8] . Com efeito, a suspeita de Woolfolk e seus colegas era que a teoria (presumivelmente tácita) da responsabilidade das pessoas mostrava-se compatibilista.

Para testá-lo, os participantes foram convidados a ler uma história sobre um agente que foi forçado por um grupo de sequestradores armados a matar um homem que estava tendo um caso com sua esposa. Na condição de "baixa identificação", o homem foi descrito como alguém horrorizado por ser forçado a matar o amante de sua esposa, mostrando não querer fazê-lo. Na condição de "alta identificação", o homem é descrito como acolhendo a oportunidade e querendo matar o amante de sua esposa. Em ambos os casos, o homem não teve escolha e matou o amante de sua esposa.

---

[8] "Identificação" é uma noção problemática, carregada de mais complexidade filosófica do que ela pode ser operacionalizada no trabalho empírico. Woolfolk et al. tentou retratar comportamentos que poderiam adequadamente ser interpretados como manifestação de identificação, ou de sua falta, sem solucionar as questões conceituais. Para alguma discussão sobre as dificuldades filosóficas não resolvidas em torno da noção de identificação, confira os trabalhos de Velleman (1992), Bratman (1996) e Watson (1996).

Segundo o esperado pela hipótese de Woolfolk e seus colegas, os participantes julgaram que o agente de alta identificação seria mais responsável, mais apropriadamente culpado e mais adequadamente sujeito à culpa do que o agente de baixa identificação[9] . Esse padrão nos juízos morais populares parece sugerir que os participantes não foram consistentemente incompatibilistas em suas atribuições de responsabilidade, porque a falta de alternativas disponíveis para o agente não foi suficiente para descartar tais atribuições.

Em resposta a tais resultados, aqueles que assumem a moralidade popular como incompatibilista podem rapidamente objetar que o estudo apenas sugere que as atribuições de responsabilidade são influenciadas pela identificação, mas nada diz sobre compromissos incompatibilistas ou sua ausência. Os participantes poderiam ainda ter acreditado haver a possibilidade de o agente fazer o contrário. Para resolver essa inquietação, Woolfolk e seus colegas também conduziram uma versão do estudo em que o homem agiu sob a influência de uma "droga de adesão". Nesse caso, os participantes estavam marcadamente menos propensos a assumirem que o homem "era livre para agir diferente do que ele fez", e mesmo assim assumiram que o agente identificado com a ação era mais responsável do que o agente que não se identificava. Tais resultados parecem representar um claro desafio à visão de que as pessoas comuns são tipicamente incompatibilistas.

Um padrão de respostas semelhante foi obtido por Nahmias, Morris, Nadelhoffer e Turner (2009), que descreveram agentes que, ao contrário, executavam comportamentos imorais em um "mundo determinista" do tipo geralmente descrito em salas de aula de filosofia. Uma variação pode ser lida como segue:

> Imagine que no próximo século descobriremos todas as leis da natureza, e construiremos um supercomputador que possa deduzir dessas leis da natureza e do estado atual de tudo no mundo exatamente aquilo que acontecerá no mundo em qualquer momento futuro. Ele poderá ver tudo sobre

---

[9] O efeito parece ser facilmente replicável. Por exemplo, o grupo Woolfolk (materiais não publicados) obteve resultados consistentes em variações relacionadas ao comportamento de valência positiva da doação de rim e ao comportamento de valência negativa de uma atrocidade cometida por um soldado obedecendo ordens.

> como o mundo é, e prever tudo sobre como ele será com 100% de precisão. Suponha que tal supercomputador existisse e olhe para o estado do universo em um determinado momento em 25 de março de 2150 d.C., vinte anos antes do nascimento de Jeremy Hall. O computador então deduz dessa informação e das leis da natureza o fato de que Jeremy decididamente roubará o Fidelity Bank às 18h, do dia 26 de janeiro de 2195. Como de costume, a previsão do supercomputador está correta; Jeremy rouba o Fidelity Bank às 18h, do dia 26 de janeiro de 2195.

Os participantes, então, foram questionados se Jeremy seria moralmente culpável. A maioria afirmou que sim, indicando terem a opinião de que um agente pode ser moralmente culpável mesmo que seus comportamentos sejam inteiramente determinados por leis da natureza. Em acordo com os resultados de Woolfolk *et al*., parece que os julgamentos dos participantes, pelo menos os que têm a ver com culpabilidade moral, não se mostraram orientados por um compromisso com o incompatibilismo.

Essa aparente indicação, no entanto, foi problematizada por Nichols e Knobe (2007), ao argumentarem que respostas ostensivamente compatibilistas traduziam erros de desempenho ocasionados por uma resposta afetiva às ações imorais dos agentes. Com o intuito de demonstrá-lo, convidaram todos os participantes a imaginar dois universos – um deles completamente governado por leis deterministas (Universo A) e outro (Universo B) em que tudo é determinado exceto as decisões humanas, que se mostram não completamente condicionadas por leis deterministas e pelo que ocorreu no passado. No Universo B, mas não em A, "cada decisão humana **não tem de acontecer** da maneira como acontece". Alguns participantes foram fixados a uma situação concreta e instados a fazerem um julgamento sobre um indivíduo específico em circunstâncias específicas, enquanto outros foram fixados a uma situação abstrata e instados a fazerem um julgamento mais geral, desvinculada de qualquer indivíduo em particular. Tinha por hipótese que a diferença entre as duas situações geraria respostas diferentes em face da relação entre determinismo e responsabilidade moral. Os participantes fixados à situação concreta leram a história de "Bill" – um homem que, agindo em um universo determinista, assassina sua esposa e filhos de maneira particularmente pavorosa – após o que são interrogados se Bill seria moralmente responsável pelo que fizera. Por outro lado, os participantes fixados à situação abstrata receberam a seguinte questão:

"é plenamente possível que uma pessoa, no Universo A, seja moralmente responsável por suas ações?" Setenta e dois por cento dos participantes vinculados à situação concreta deram uma resposta compatibilista, responsabilizando Bill no Universo A, enquanto menos de quinze por cento dos participantes vinculados à situação abstrata deram uma resposta compatibilista, acreditando plenamente possível que as pessoas sejam moralmente responsabilizadas no Universo A determinista.

De acordo com trabalhos experimentais anteriores, demonstrando que o aumento da excitação afetiva intensificava as respostas punitivas à infração (LERNER; GOLDBERG; TETLOCK, 1998), Nichols e Knobe levantaram a hipótese de que aquelas respostas compatibilistas observadas acima resultariam de uma natureza afetivamente saturada com materiais de estímulo. Sendo esse elemento afetivo eliminado dos materiais (como ocorre na situação abstrata), os participantes de fato apresentarão um padrão incompatibilista de respostas.

Recentemente, a linha de raciocínio de Nichols e Knobe sofreu críticas, sustentadas em duas direções. Na primeira delas, diversos estudos tentaram tornar sistematicamente manipulável quanto o comportamento imoral efetivado poderia ser provocante afetivamente, sem descobrir se essas mudanças alteram significativamente os juízos de responsabilidade moral dos participantes em cenários determinísticos. Em vez disso, tais diferenças parecem simplesmente ser melhor explicadas pelo fato de o caso ter sido descrito abstrata ou concretamente (*vide* COVA *et al*., 2012, que trabalha com pacientes que possuem demência frontotemporal[10] , e FELTZ; COVA, 2014, para uma meta-análise). No segundo caso, uma linha de estudos separada por Murray e Nahmias (2014) argumentou que os participantes que apresentaram um padrão de resposta aparentemente incompatibilista estavam incorrendo em um erro crítico acerca de como entendiam o cenário determinístico. Especificamente, eles argumentam que os participantes assumiram erroneamente, em cenários determinísticos, os agentes ou seus estados mentais como sendo "ignorados" na cadeia causal que tornou possível o seu comportamento. Para fundamentar esse argumento, Murray e Nahmias (2014) demonstraram que, no caso em que as análises se restringiam aos participantes que claramente não

---

[10] N.T.: A demência frontotemporal, antes conhecida como *doença de Pick*, caracteriza-se por um conjunto de distúrbios que atingem os lobos frontais do cérebro, ocasionando alterações de comportamento e personalidade, bem como ineficiência na compreensão e na produção da fala.

assumiam o agente como ignorado, eles o julgavam como moralmente responsável (condenável, etc.) apesar de situado em um universo determinista. Não surpreende que essa linha de argumento, por sua vez, tenha inspirado várias outras respostas contrárias, tanto empíricas (ROSE; NICHOLS, 2013) quanto teóricas (BJÖRNSSON; PEREBOOM, 2016), que alertam contra essas conclusões de Murray e Nahmias.

Enquanto o debate prossegue para definir qual posição, compatibilista ou incompatibilista, melhor captura os juízos morais populares de agentes em universos determinísticos, uma linha paralela de pesquisa surgiu interessada no que pode ser amplamente considerada a forma contemporânea mais convincente de argumento para o incompatibilismo: argumentos de manipulação (*vide* MELE, 2006, 2013; PEREBOOM, 2001, 2014). A versão Quatro-Casos de Pereboom, por exemplo, começa descrevendo o caso de um agente chamado Plum que é manipulado por neurocientistas que se valem de uma tecnologia similar ao rádio para modificar estados neurais de Plum, o que termina por fazê-lo querer e então decidir matar um homem chamado White. Neste caso, parece claro que Plum não decidiu livremente matar White. É possível comparar esse caso a um segundo, no qual a equipe de neurocientistas programou Plum desde o início de sua vida de modo a fazê-lo desenvolver o desejo (levando-o a tomar a decisão) de matar White. O incompatibilista argumenta que os dois casos não diferem, de forma relevante, para fazer Plum agir livremente e, com isso, mais uma vez, parece que Plum não decidiu livremente matar White. Agora, compare os dois casos com um terceiro, no qual o desejo e a decisão de Plum, de matar White, foram determinados pelo seu meio cultural e social, em vez de uma equipe de neurocientistas. Uma vez que a única diferença entre o segundo e o terceiro caso é o processo tecnológico particular através do qual os estados mentais de Plum foram determinados, ele outra vez parece não ter decidido livremente matar White. Por fim, em um quarto e último caso, o desejo e a decisão de Plum de matar White foram conjuntamente determinados pelos estados passados e pelas leis da natureza em nosso próprio universo determinístico. Com relação a esses quatro casos, Pereboom argumenta que, não havendo diferença alguma entre esses quatro casos que seja relevante ao livre-arbítrio, se Plum não seria moralmente responsável no primeiro caso, então não o seria também no quarto.

Como resposta a esse tipo de argumento para um incompatibilismo baseado em manipulação, vários pesquisadores tiveram por objetivo esboçar melhor um quadro empírico destinado aos juízos morais cotidianos acerca de agentes manipulados. Tal linha de investigação tem se mostrado produtiva em dois níveis. Em primeiro lugar, um número crescente de estudos empíricos tem investigado os

juízos de responsabilidade moral, relativos a casos de manipulação, que fornecem agora uma imagem psicológica mais clara do porquê os agentes manipulados são julgados como não possuindo livre arbítrio e responsabilidade moral. Em segundo lugar, o trabalho teórico contínuo, esclarecido por esse quadro empírico, levanta novas razões que põem em dúvida se, afinal, os argumentos baseados em manipulação realmente fornecem evidências contra o compatibilismo.

Uma linha de pesquisa empírica, liderada por Chandra Sripada (2012), tem examinado se os agentes manipulados são percebidos como não-livres em razão de (a) lhes faltar o controle último sobre suas ações (capacidade considerada pelos incompatibilistas como essencial à responsabilidade moral) ou em razão de (b) terem prejudicadas suas capacidades psicológicas ou volitivas (aquelas enfocadas pelos compatibilistas). Usando uma abordagem estatística chamada Modelagem de Equações Estruturais (ou SEM)[11], Sripada descobriu que os juízos de responsabilidade moral dos participantes eram melhor explicados se eles acreditassem que as capacidades psicológicas e volitivas do agente tinham sido prejudicadas pela manipulação, e não por reconhecerem o agente como não tendo controle sobre suas ações. Tal descoberta sugere que os padrões de julgamento, em casos de manipulação, são mais consistentes com os prognósticos do compatibilismo do que com os do incompatibilismo.

Seguindo uma abordagem diferente, Phillips e Shaw (2014) demonstraram que a redução da responsabilidade moral, tipicamente observada em casos de manipulação, dependeria criticamente do papel de um manipulador **intencional**. Em especial, as pessoas comuns mostraram-se capazes de distinguir entre (1) a responsabilidade moral dos agentes que são compelidos a realizar um determinado ato, devido às características da situação em que se encontram (ou seja, um determinismo situacional), e (2) a responsabilidade moral dos agentes que são compelidos, intencionalmente por outro agente, a fazer o mesmo ato (ou seja, uma manipulação). Esta abordagem sugere que a prática cotidiana de aferir liberdade e responsabilidade mostra-se claramente capaz de distinguir entre casos que envolvem e que não envolvem um manipulador intervindo com o propósito de fazer com que o agente manipulado pratique uma ação imoral. Uma série de estudos de Murray e Lombrozo (2016) fazem avançar tais descobertas ao fornecerem evidências

---

[11] N.T.: Tradução em língua portuguesa para *Structural Equation Modeling.*

de que a redução específica da responsabilidade moral, como resultado de alguém ter sido manipulado, é proveniente da percepção de que os estados mentais do agente foram ignorados.

Em geral, duas lições decorrem desse estudo sobre a prática cotidiana de avaliar a responsabilidade moral de agentes manipulados: (1) a moralidade popular ilustra um modo natural para distinguir entre os diferentes casos levados em conta nos argumentos baseados em manipulação (os que envolvem a intervenção intencional de um manipulador X os que não o fazem) e (2) a moralidade popular traça uma conexão íntima entre a responsabilidade moral de um agente e as suas capacidades mentais e volitivas. Fundado em um quadro empírico cada vez mais claro, Deery e Nahmias (2017) formalizaram esses princípios básicos em um trabalho teórico cuja argumentação pretende estabelecer uma maneira arquetípica de distinguir entre a responsabilidade moral de agentes determinados e a de agentes manipulados.

Ainda que, atualmente, a maior parte das evidências seja a favor da visão de que a moralidade popular adere a um tipo de "compatibilismo natural" (COVA; KITANO, 2013), este permanece sendo um tópico controverso, e novos trabalhos surgem continuamente, posicionados em ambos os lados do debate (ANDOW, COVA, 2016; BEAR, KNOBE, 2016; BJÖRNSSON, 2014; FELTZ, MILLAN, 2013; FIGDOR, PHELAN, 2015; KNOBE, 2014). Ambos os lados, contudo, parecem estar de acordo quanto a um ponto importante: a posição crítica em favor de estudos empíricos cuidadosos (BJÖRNSSON, PEREBOOM, 2016; KNOBE, 2014; NAHMIAS, 2011).

4. Ética da virtude e ceticismo sobre o caráter

A partir disso, as abordagens da psicologia moral informadas empiricamente têm se mostrado mais proeminentes nas discussões acerca do caráter moral e da virtude. Eis o resultado de décadas de experimentação em psicologia social "situacionista": os aspectos que até então pareciam discretos em certas circunstâncias têm sido com frequência denunciados como capazes de impactar o comportamento moral, de modo só aparentemente arbitrário, embora algumas vezes alarmante. Entre as descobertas que mais tem interessado os filósofos, destacam-se:

- Estudo da Cabine Telefônica (Isen, Levin, 1972, p.387): pessoas que tinham acabado de encontrar um centavo no retorno de moedas de um telefone público eram 22 vezes

mais propensas a ajudar uma mulher que havia deixado cair alguns papéis do que aquelas que não encontraram nenhum centavo (88% contra 4%).

- Estudo do Bom Samaritano (Darley, Batson, 1973, p.105): transeuntes morosos eram 6 vezes mais propensos a ajudar um infeliz que parecia estar em um sofrimento expressivo do que transeuntes apressados (63% contra 10%).

- Experimentos de Obediência (Milgram, 1974): participantes puniram repetidamente uma vítima, que gritava após receber choques elétricos realistas (ainda que simulados), seguindo o pedido educado de um especialista[12] .

- Estudo da Prisão de Stanford (Zimbardo, 2007): estudantes universitários, ao interpretarem o papel de "guardas" em uma prisão simulada, submetiam estudantes "prisioneiros" a grotescos abusos verbais e emocionais.

Esses experimentos fazem parte de **uma extensa literatura empírica**, em que psicólogos sociais continuam a descobrir que, uma vez mais, omissões decepcionantes e ações terríveis são facilmente induzidas por aspectos situacionais aparentemente insuspeitos[13] . O que surpreende não é o fato de as pessoas falharem quanto aos padrões de uma boa conduta, mas poderem ser **facilmente** induzidas a fazê-lo.

Ao explorarem essa observação, os ditos "céticos acerca do caráter" costumam afirmar que, se a conduta moral varia de modo acentuado, muitas vezes para pior, em decorrência de pequenas perturbações nas circunstâncias, o caráter ostensivamente bom acaba por fornecer uma garantia bastante limitada para uma boa conduta. Em acréscimo a essa constatação em *psicologia descritiva*, referente

---

[12] N.T.: O filme *Experimenter* (2015) conta a história do psicólogo Stanley Milgram, que se tornou conhecido após o resultado surpreendente de seu experimento.

[13] Para um exame da literatura empírica, confira os trabalhos de Doris (2002), Miller (2013, 2014) e Ross e Nisbett (1991).

à fragilidade do caráter moral, alguns desses céticos avançam também uma tese em *ética normativa*, a de que o caráter deveria merecer menos atenção do que aquela que tradicionalmente ele recebe no interior da reflexão ética[14] .

O ceticismo quanto ao caráter antepõe-se à influente aplicação da *ética da virtude* na contemporaneidade – sustentada pelo avanço da teoria ética como requerendo maior atenção ao caráter –, o que exige dos eticistas da virtude lhe oferecerem uma resistência vigorosa[15] . A discussão, quase sempre inflamada, fez surgir uma grande literatura, afeita a certo campo de "estudos de caráter" efusivamente interdisciplinar (*vide* MILLER *et al*. 2015)[16] . A literatura é extensa demais para os limites deste verbete, mas nos esforçaremos para delinear algumas das suas principais temáticas.

Em primeiro lugar, cabe observar que a própria ciência, que inspira os céticos quanto ao caráter, pode ela mesma se mostrar sujeita ao ceticismo. Dada a história desigual das ciências humanas, é possível argumentar que as descobertas relevantes são demasiado incertas para servirem como restrição à teorização filosófica. Tal contestação acaba sendo potencialmente apoiada pelas falhas recentes nas respostas proeminentes em psicologia social.

A psicologia em questão, assim como grande parte da ciência, é um ofício inacabado. No entanto, alegar respostas controversas, e a concomitante suspeita sobre a ciência, é insuficiente para descartar de imediato a psicologia. Conclusões filosóficas não deveriam ser baseadas em estudos escassos; a tarefa do consumidor

---

[14] Sobre o ceticismo acerca do caráter em filosofia, confira Doris (1998, 2002, 2005, 2006, 2010, no prelo) e Doris e Stich (2005). Para posições relacionadas, Alfano (2013), Harman (1999, 2000), Machery (2010), Merritt (2000, 2009), Meritt, Doris e Harman (2010) e Vranas (2005). Em psicologia, Mischel (1968) e Ross e Nisbett (1991). Uma variedade de perspectivas posteriores foi coletada em Donnellan et al. (2009).

[15] Sobre as respostas ao ceticismo quanto ao caráter, confira Adams (2006), Annas (2005), Appiah (2008, p. 33-72), Arpaly (2005), Badhwar (2009), Besser-Jones (2008), Flanagan (2009, p. 54-56), Kamtekar (2004), Knobe, Leiter (2007), Kupperman (2001), Miller (2003), Montmarquet (2003), Prinz (2009); Russell (2009, p. 227-331), Samuels, Casebeer (2005), Sarkissian (2010), Solomon (2003, 2005), Snow (2010), Sreenivasan (2002), Swanton (2003, p. 30-31), Upton (2009) e Webber (2006a, 2006b, 2007a, 2007b).

[16] Para uma bibliografia mais extensa, consulte *Skepticism about Character,* de Mark Alfano, disponível em: http://philpapers.org/browse/skepticism-about-character. Acesso em: 02 jul. 2022.

filosófico de ciência é identificar **tendências** naqueles vínculos de evidência **convergentes** (DORIS, 2015, P. 49; 56; MACHERY, DORIS, no prelo). A observação que motiva o ceticismo acerca do caráter – a espantosa vulnerabilidade situacional dos comportamentos, indicada acima – vê-se apoiada em uma ampla variedade de descobertas científicas, tanto quanto na presença de temas recorrentes em história e biografias (DORIS, 2002, 2005). A forte determinabilidade situacional dos comportamentos é aceita como fato por um número elevado de cientistas envolvidos na pesquisa, o que não o torna, por isso, muito contestado em debates sobre o ceticismo quanto ao caráter.

Contudo, as implicações filosóficas desse fato, após considerável debate, permanecem um tema controverso. As várias respostas ao ceticismo quanto ao caráter não precisam ser encaminhadas separadamente, porque algumas delas podem ser combinadas como parte de uma defesa versátil. Réplicas diferentes têm diferentes pontos fortes e fracos, especificamente quanto às distintas evidências nas quais os céticos confiam. Já que os fenômenos não são unitários, acomodá-los todos pode impossibilitar uma resposta unitária.

Uma maneira de neutralizar o ceticismo empiricamente motivado – apelidado por Alfano (2013) de "esquiva" – está em simplesmente negar que a ética da virtude faz afirmações empíricas. Segundo esse entendimento, a ética da virtude se deixa moldar por um empreendimento "puramente normativo", cujo propósito está em erguer ideais éticos numa completa ausência de compromissos empíricos relativos às questões psicológicas reais dos humanos. Tal tipo de pureza talvez seja menos honrosa do que aquela honrada na infração: historicamente, a ética da virtude vem sendo tipificada pelo interesse em como as pessoas reais *se* **tornam** boas. Aristóteles (*Ética a Nicômaco*, 1099b 18-19) pensava que qualquer pessoa não "mutilada", no que diz respeito à capacidade para a virtude, pode adquiri-la "através de certo tipo de estudo e atenção", e os aristotélicos contemporâneos procuram enfatizar a importância de uma educação enquanto desenvolvimento moral (*vide*, ANNAS, 2011). De modo geral, abordagens com base na virtude têm sido reivindicadas por serem mais vantajosas do que seus principais concorrentes kantianos e consequencialistas, sobretudo quanto ao "realismo psicológico" – a vantagem de uma psicologia moral mais autêntica (*vide* ANSCOMBE, 1958, P. 1, P. 15; WILLIAMS, 1985; FLANAGAN, 1991, P. 182; HURSTHOUSE, 1999, p. 19-20).

Com razão, o afastamento de um compromisso empírico permite que a ética da virtude escape da ameaça empírica: obviamente, uma evidência empírica não pode ser utilizada para debilitar uma teoria que **não** faz afirmações empíricas.

Não parece claro, contudo, que tais teorias possam sustentar vantagens tradicionalmente reivindicadas para as teorias da virtude em relação ao desenvolvimento moral e ao realismo psicológico. De qualquer modo, elas não expõem contribuições para a psicologia moral empírica e, portanto, não precisam ser discutidos aqui.

Antes de observar como o debate em psicologia moral pode avançar, é necessário corrigir uma caracterização equivocada que tem servido para impedir esse progresso. É muitas vezes dito, em referência particularmente a Doris (1998, 2002) e Harman (1999, 2000), que o ceticismo sobre o caráter chega ao ponto de afirmar os traços de caráter como "não existentes" (*vide* Flanagan, 2009, p. 55). Frequentemente, essa atribuição é feita sem qualquer fundamento, e quando fornecem a documentação, ela é indicada geralmente com referência a certas observações antigas, apontadas de modo peculiar por Harman (1999). Contudo, numa contribuição mais recente, Harman (2009, p. 241) pôde afirmar: "não penso que a psicologia social demonstre não haver traços de caráter". De sua parte, Doris tem repetidamente afirmado que os traços de caráter existem, e com frequência chama nossa atenção para essas afirmações (DORIS, 1998, P. 507-509; 2002, P. 62-66; 2005, P. 667; 2010, P. 138-141; DORIS, STICH, 2005, P. 119-120; DORIS, PRINZ, 2009).

A bem da verdade, dizer que "traços [de caráter] não existem" equivale a negar que diferenças de disposição individuais existem – uma perspectiva improvável que céticos e anticéticos parecem unidos em rejeitar. Não surpreende que tal visão improvável seja seriamente subscrita tanto em filosofia quanto em psicologia, e que jamais seja endossada quer pelos mais agressivos críticos da personalidade, aqueles situacionistas da psicologia social como Ross e Nisbett (1991), quer pelo próprio santo padroeiro do situacionismo em psicologia da personalidade: Mischel (1999, p. 45). Se Mischel rejeita uma abordagem baseada nos traços de caráter, seu ceticismo é referido não às diferenças individuais da disposição em geral, mas a certa *abordagem particular aos traços*.

Desse modo, a questão acerca da existência ou não dos traços de caráter não é *o* **problema** que de maneira enfática rivaliza as abordagens mais e menos céticas quanto ao caráter. Hoje, todas as principais correntes que participam desse debate são "interacionistas", ao tratarem resultados comportamentais como função de uma pessoa (complexa) em interação com certa situação (MEHL *et al*. 2015) – é provável que a maioria dos participantes tenha sido sempre considerada assim (DORIS, 2002, p. 25-26). Os atuais programas de pesquisa em personalidade e psicologia social aplicam livremente variáveis **tanto** pessoais **quanto** situacionais

(*vide* CAMERON, PAYNE, DORIS, 2013; LEIKAS, LÖNNQVIST, VERKASALO, 2012; SHERMAN, NAVE, FUNDER, 2010). O problema que vale a discussão não é saber se existem diferenças individuais da disposição geral, mas saber **como essas diferenças deveriam ser caracterizadas** e como (ou se) essas diferenças individuais, quando caracterizadas apropriadamente, **deveriam determinar o pensamento ético**.

Um aspecto importante das primeiras incursões no âmago do ceticismo esteve em mostrar que os céticos tendiam a realçar as *implicações comportamentais* dos traços de caráter, em vez de enfocar os *antecedentes psicológicos* do comportamento (DORIS, 2015, p. 15). Defensores da ética da virtude observam que os céticos quanto ao caráter têm muito a dizer sobre a implicação da variação situacional no comportamento, e pouco a dizer sobre os processos psicológicos a ela subjacentes, do que resulta uma negligência à ordem racional na vida das pessoas (ADAMS, 2006, p. 115-232). Esses eticistas da virtude sustentam que a variação comportamental, sendo provocativa ao ceticismo, evidencia não uma falta de confiabilidade, mas de sensibilidade racionalmente apropriada para diferentes situações (ADAMS, 2006; KAMTEKAR, 2004). A pessoa virtuosa, tal como o modelo aristotélico do *phronimos* ("o homem de sabedoria prática"), pode ser às vezes probo, às vezes dissimulado, ou às vezes brigar e às vezes fugir, a depender das exigências éticas particulares de cada circunstância.

Por exemplo, no chamado Estudo do Bom Samaritano, os transeuntes apressados dirigiam-se a um compromisso no qual haviam concordado em fazer uma exposição. Essas pessoas, talvez, tenham feito uma determinação racional — talvez mesmo uma determinação eticamente defensável — de avaliar exigências de pontualidade e profissionalismo acima da necessidade ética de verificar o bem-estar de um estranho em um sofrimento aparente. Por mais atrativa que pareça a alguém tal explicação para esse caso (vale notar que alguns transeuntes apressados de Darley e Batson [1973] sequer perceberam a vítima, o que compromete as explicações em termos de suas discriminações racionais), há outros em que a "resposta racional" claramente se manifesta como pouco atraente. Seriam os casos de influências eticamente irrelevantes (ver acima, seção 2; DORIS, STICH, 2005), nos quais parece improvável que a influência possa ser mencionada como parte de uma explicação racionalizada do comportamento. É de todo modo curiosa a alegação de que não encontrar um centavo serve como **justificativa** para a omissão em ajudar — ou, no que interessa aqui, encontrar um centavo serve como **justificativa** para fazê-lo.

Faz-se certamente apropriado para os éticos da virtude enfatizarem a racionalidade prática em suas explicações acerca do caráter. É de fato um tema central na tradição que remonta ao próprio Aristóteles, provavelmente o filósofo canônico mais citado na ética das virtudes contemporânea. Só que, embora a resposta racional possa acomodar inicialmente algumas das mais incômodas evidências comportamentais, ela esbarra em muitas dificuldades empíricas. Existe uma extensa literatura empírica que problematiza concepções banais sobre a racionalidade: psicólogos vêm documentando interminavelmente uma variedade desalentadora de erros de raciocínio (BARON, 1994, 2001; GILOVICH *et al*., 2002; KAHNEMAN *et al*., 1982; TVERSKY; KAHNEMAN, 1973; KRUGER; DUNNING, 1999; NISBETT; BORGIDA, 1975; NISBETT; ROSS, 1980; STICH, 1990; TVERSKY; KAHNEMAN, 1981). À luz dessas evidências, os céticos quanto ao caráter denunciam os caprichos que afligem o comportamento como também afligindo o raciocínio (ALFANO, 2013; OLIN; DORIS, 2014).

A pesquisa que fundamenta essa avaliação desalentadora da racionalidade humana é controversa, e nem todos os psicólogos pensam ser as coisas assim tão sombrias (GIGERENZER, 2000; GIGERENZER *et al*., 1999), para comentários filosóficos, confira os trabalhos de Samuels e Stich (2002). No entanto, se a ética da virtude deve possuir uma psicologia moral empiricamente crível, ela precisa dar conta dos desafios empíricos ao raciocínio prático: afinal, como a excelência pertinente ao raciocínio prático pode ser desenvolvida?

Frente ao desafio posto à racionalidade prática, eticistas da virtude podem responder que suas teorias se preocupam com o raciocínio **excelente**, não com o raciocínio **cotidiano** estudado em psicologia. Da sabedoria prática, e da virtude ética que ela sustenta, espera-se que sejam **raras**, em vez de amplamente instanciadas. Um tal estado de coisas, costuma-se dizer, é bastante compatível com o comportamento perturbador, ainda que não excepcionalmente, demonstrado em experimentos como o de Milgram (*vide* ATHANASSOULIS, 1999, P. 217-219; DEPAUL, 1999; KUPPERMAN, 2001, p. 242-243). Se essa consideração deve fazer parte de uma psicologia moral robustecida empiricamente, sem acabar repleta de especulações não verificadas, faz-se necessária uma explicação detalhada e empiricamente sustentada de como esses poucos virtuosos chegam a sê-lo – vale lembrar que uma ênfase no desenvolvimento moral é central para a tradição da ética da virtude. Além disso, se a ética da virtude deve possuir amplas implicações práticas – por oposição a ser vista meramente como celebração de uma minúscula "elite da virtude" – ela deveria oferecer uma explicação de como os muitos-pouco-

virtuosos podem ao menos **se aproximar** satisfatoriamente da virtude.

Esse ponto é sublinhado pelo fato de que, para algumas das evidências incômodas, como as do Estudo da Prisão de Stanford, a preocupação não está em saber sobretudo se as pessoas falham quanto a padrões de virtude, e sim que elas falhem quanto a padrões *mínimos de decência*. Por certo, uma abordagem ética que louve o desenvolvimento moral, mesmo reconhecendo (ou melhor, insistindo) que a maioria das pessoas não atingirá esse ideal, poderia oferecer uma explicação de como as pessoas podem se tornar minimamente decentes.

Nos últimos tempos, os proponentes da ética da virtude pouco a pouco vêm desenvolvendo uma sugestiva solução para esse problema: virtude é uma habilidade adquirida através de certa prática esforçada, ou seja, ela é um tipo de perícia (Annas, 2011; Bloomfield, 2000, 2001, 2014; Jacobson, 2005; Russell, 2015; Snow, 2010; Sosa, 2009; Stichter, 2007, 2011). Quer dizer, os virtuosos são peritos em moralidade e, dada a associação aristotélica entre virtude e felicidade, especialistas na vida.

Uma literatura científica extensa tem indicado que o desenvolvimento de habilidades especializadas requer ampla preparação, seja o praticante um romancista, um médico ou um mestre de xadrez – cerca de 10.000 horas de "prática intencional", segundo uma generalização popular (ERICSSON, 2014; ERICSSON *et al*., 1993). A "regra das 10.000 horas" provavelmente exagera na simplificação, porém não restam dúvidas de que a obtenção de experiência requer um treinamento intenso. Por conta disso, as pessoas raramente alcançam projeção em mais de uma área: especialistas em "trivia de beisebol"[17] , por exemplo, exibem memória superior para um material relacionado ao beisebol, sem exibirem a mesma qualidade para um material estranho ao beisebol (CHIESI *et al*., 1979). Do mesmo modo, tornar-se um perito em moralidade, ou (ainda mais ambiciosamente) um especialista em vida no geral, exigiria aparentemente uma forma **altamente** generalizada de especialização: para ser bom, haveria **muito** em que se tornar bom. Além disso, não parece muito claro o que está envolvido nessa prática intencional quanto à vida: como exatamente alguém se torna melhor em ser bom?

---

[17] N.T.: Trívia é o termo que designa os detalhes ou curiosidades relativas a algo. Os chamados *sport trivia questions* são testes de memória sobre diversos assuntos relativos a modalidades esportivas, espécie de Quiz ou *Puzzle*, entre os quais o de Beisebol é famoso.

Uma questão que parece óbvia está em saber sobre a especificidade do "bom" em causa. Perícias, como a do xadrez, têm efetivamente sido estudadas, em parte, por haver certos padrões aceitos de excelência, a pontuação "ELO" usada para classificar os jogadores de xadrez (GLICKMAN, 1995). Pode-se dizer com alegria que em xadrez não há céticos. Mas, historicamente, tem havido **muitos** céticos morais. Contudo, se não há conhecimento moral, como pode haver peritos morais? E mesmo que existam peritos morais, há o problema de saber como eles podem ser identificados, uma vez que não é claro se possuímos algum padrão que seja independente da própria opinião desses peritos (algo similar a saber como ganhar partidas de xadrez) para podermos identificá-los. Sobre a **metaética da especialização**, confira Mcgrath (2008, 2011).

Ainda que essas notórias dificuldades filosóficas possam ser resolvidas – como pensam poder resolvê-las os defensores da abordagem da virtude como perícia –, certos tópicos continuam complicados, porque se a perícia moral é como as outras, então apenas a prática – supondo que tenhamos uma noção clara do que implica uma "prática moral"— seria insuficiente. Mesmo que a prática influa na obtenção de uma perícia, outros fatores, como o talento, também são importantes (HAMBRICK *et al*., 2014; MACNAMARA *et al*., 2014). Além disso, alguns dos recursos necessários podem estar distribuídos de forma bastante desigual entre a população: apenas a prática não pode fazer de um jóquei um atacante da NFL, ou de um atacante da NFL um jóquei.

Quais seriam então os recursos naturais exigidos para uma perícia moral e quão amplamente eles se fazem distribuídos entre a população? Se forem raros, como as habilidades de um mestre de xadrez ou a força de um jogador de linha da NFL, a virtude também será rara. Alguns eticistas da virtude acreditam que ela deva ser amplamente alcançável, e eles resistirão à conclusão acima, confira Adams (2006, p. 119-123) e, provavelmente, Aristóteles, em *Ética a Nicômaco* (1099b 15-20). Todavia, mesmo aqueles que abraçam a raridade da virtude exigem levar em conta quais seriam os recursos naturais necessários. Caso eles anseiem também considerar como os menos bem-dotados podem no mais alcançar um nível mínimo de decência, eles deveriam ter algo a dizer sobre como o desenvolvimento moral ocorrerá em uma população dotada de recursos amplamente variados.

O que se faz necessário, para fazer avançar o estudo sobre as pesquisas quanto ao caráter moral, encontra-se na consideração dos fatores biológicos, psicológicos e sociais que seriam exigidos para o desenvolvimento moral bem-sucedido – no modelo da perícia, as condições que conduzem ao desenvolvimento

da “habilidade moral”. Isso, obviamente, é uma empreitada importante, e a pesquisa necessária para abordar sistematicamente essa questão pode ser dita estar ainda em um estágio primário. No entanto, o modelo de perícia para a virtude, ao explorar vinculações com áreas nas quais a aquisição de habilidades já foi bem estudada, como em música e esporte, pode fornecer um enquadramento capaz de mover a discussão sobre o caráter para além das conjecturas e suposições, pouco apoiadas empiricamente, sobre “habituação” que têm sido muito frequentemente discutidas na literatura anterior (DORIS, 2015, p. 128).

5. Egoísmo *versus* Altruísmo

As pessoas costumam se comportar de modo a beneficiar os outros, e por vezes o fazem mesmo sabendo que o resultado terá um custo alto, desagradável ou perigoso. Entretanto, ao menos desde a clássica discussão realizada por Platão no segundo livro de *A República*, o debate acerca do **porquê** as pessoas se comportam dessa maneira se acirrou. Seus motivos seriam altruístas, ou seu comportamento é motivado, em última análise, por um interesse próprio? Eis a conhecida resposta de Hobbes:

> Ninguém oferece nada senão com a intenção de obter um bem para si mesmo: isso porque a dádiva é voluntária; e todos os atos voluntários têm por objetivo, para cada homem, seu próprio bem; do qual, percebendo os homens que ele será frustrado, nenhum princípio de benevolência ou confiança haverá, nem, consequentemente, de qualquer ajuda mútua (1981)[18].

Visões como a de Hobbes passaram a ser chamadas de *egoísmo*[19], e mesmo uma concepção tão deprimente da motivação humana como essa teve

---

[18] N.T.: No capítulo 15 da primeira edição, lançado em 1651.

[19] Os termos “egoísmo” e “altruísmo” estão sendo usados aqui referindo-se a visões descritivas sobre a natureza da motivação humana. Alguns preferem se referir a tais visões através das expressões “egoísmo psicológico” e “altruísmo psicológico”, para distingui-las do “egoísmo ético” e “altruísmo ético”, que são visões prescritivas sobre como as pessoas devem se comportar.

inúmeros defensores filosóficos eminentes, incluindo Bentham, J.S. Mill e Nietzsche[20] . Vozes dissidentes, ainda que talvez em menor número, não se mostraram menos ilustres. Butler, Hume, Rousseau e Adam Smith argumentaram que, pelo menos às vezes, a motivação humana é genuinamente altruísta.

Embora a questão da motivação humana, que antepõe explicações egoístas e altruístas, seja amplamente empírica, é fácil perceber por qual razão os filósofos pensam que essas respostas concorrentes expõem consequências importantes para a teoria moral – por exemplo, o famoso argumento de Kant, de que uma pessoa deve agir "não por inclinação, mas por dever, pois assim sua conduta adquire, desde logo, verdadeiro valor moral" (1949, sec. 1, §.12)[21] . No entanto, o egoísmo sustenta que *toda* motivação humana é, em último caso, autointeressada e, portanto, as pessoas **não poderiam** agir "por dever" da maneira que Kant argumenta. Desse modo, se o egoísmo é verdadeiro, então, segundo a explicação kantiana, nenhuma conduta poderia ter "verdadeiro valor moral". Além disso, se o egoísmo é verdadeiro, ele parecerá impor uma forte restrição sobre como uma teoria moral poderia responder à venerável questão: "por que eu deveria ser moral?" – já que, como Hobbes claramente observou, a resposta terá que fundamentar a motivação, para o agente ser moral, no seu interesse próprio[22] .

Enquanto o debate egoísmo X altruísmo tem despertado, historicamente, grande interesse filosófico, a problemática central permanece relativa a questões psicológicas acerca da natureza da motivação humana, de modo que não surpreende constatar a grande quantidade de pesquisas empíricas, desenvolvidas por psicólogos, cujo objetivo é determinar qual visão estaria correta. Alguns dos mais influentes e filosoficamente sofisticados trabalhos empíricos sobre essa questão foram realizados por Daniel Batson e seus colaboradores. A conclusão que Batson extrai de seu trabalho é que as pessoas se comportam, às vezes, de forma altruísta, nas quais

---

[20] A interpretação de textos de época, de fato, quase sempre é problemática. Embora existam passagens nas obras de cada um desses filósofos que podem ser interpretadas como defendendo o egoísmo psicológico, os comentadores costumam questionar se essas passagens refletem a opção considerada pelo autor.

[21] N.T.: Primeira edição, de 1785.

[22] Para outras informações sobre a história e as implicações filosóficas do debate, confira Broad (1930), Macintyre (1967), Nagel (1970), Batson (1991, cap. 1-3), Sober e Wilson (1998, cap. 9) e Dixon (2008).

a emoção da empatia desempenha um importante papel para gerar motivação altruísta[23] . Muitos não parecem convencidos. Para uma discussão dos experimentos de Batson, da conclusão que ele tira e das razões para o ceticismo quanto a essa conclusão, recomendamos as seções 5 e 6 do verbete **Abordagens Empíricas do Altruísmo**[24] . Nesta seção, vamos nos concentrar em algumas preliminares filosóficas que são necessárias antes de alguém mergulhar na literatura empírica.

Uma questão crucial que precisa ser abordada é: sobre o que exatamente é o debate; o que é altruísmo? Infelizmente, não existe uma resposta incontroversa para esta pergunta, uma vez que pesquisadores de muitas disciplinas, incluindo filosofia, biologia, psicologia, sociologia, economia, antropologia e primatologia já escreveram sobre altruísmo – e autores de diferentes disciplinas tendem a usar o termo "altruísmo" de modos bem diferentes. Mesmo entre os filósofos, o termo tem sido usado com significados valiosos, porém distintos. Contudo, há uma consideração sobre o altruísmo – composta, na verdade, por um conjunto de descrições intimamente relacionadas – que desempenha um papel central seja na filosofia, seja em grande parte da psicologia, incluindo o trabalho de Batson. Vamos chamá-la de "a consideração padrão". Esse será nosso foco no restante desta seção[25] .

De acordo com essa consideração padrão, uma ação será altruísta se for motivada pelo desejo último de bem-estar de outra pessoa. Tal formulação convida a questionar sobre (1) o que se entende por um comportamento que é **motivado pelo desejo último**, e sobre (2) a distinção entre **desejos que são por autointeresse** e **desejos que são para o bem-estar dos outros**.

Embora a segunda questão exija uma consideração cuidadosa para qualquer abordagem mais completa, alguns exemplos disponíveis, ainda que grosseiros, dessa distinção serão suficientes aqui[26] . Desejos de salvar a vida de alguém, de

---

[23] Para um apanhado do trabalho de Batson, veja Batson (1991, 2011).

[24] N.T.: Verbete da SEF, disponível em: https://plato.stanford.edu/entries/altruism-empirical/. Acesso em: 02 jul. 2022.

[25] Algumas versões da consideração padrão foram oferecidas por outros autores, incluindo Broad (1950), Feinberg (1999), Sober e Wilson (1998, cap. 6-7), Rachels (2003, cap. 6), joyce (2006, cap.1), Kitcher (2010, 2011, cap.1), May (2011a) e muitos outros. Para uma discussão de outras considerações sobre o altruísmo, consulte o verbete sobre **Abordagens empíricas ao altruísmo** (seções 2, 3 e 4), disponível em: https://plato.stanford.edu/entries/altruism-empirical/. Acesso em: 02 jul. 2022.

[26] Para alguma discussão substantiva acerca da questão, veja Cialdini *et al*. (1997), Sober e Wilson (1998, cap. 7), Stich et al. (2010), May (2011a, 2011b, 2011c) e Slote (2013).

aliviar o sofrimento alheio ou de fazer alguém feliz são casos paradigmáticos de um tipo de desejo pelo bem-estar alheio, enquanto desejos de experimentar prazer, enriquecer e tornar-se famoso são exemplos típicos de desejos autointeressados. Esses desejos autointeressados de experimentar prazer e evitar a dor têm desempenhado um papel especialmente relevante para o debate, dado que uma versão do egoísmo, muitas vezes chamada de *hedonismo*, sustenta que esses são nossos **únicos** desejos derradeiros.

Diante disso, a primeira questão, referente aos desejos últimos, requer uma exposição mais completa; esta pode ser satisfatoriamente elaborada com o auxílio de uma habitual descrição do **raciocínio prático**[27] . Segundo essa descrição, o raciocínio prático é um processo causal através do qual um desejo e uma crença dão origem ou sustentam outro desejo. Por exemplo, o desejo de beber um expresso e a crença de que o melhor lugar para tomar um expresso é no bar situado na Main Street pode causar o desejo de ir ao bar da Main Street. Esse desejo pode então unir forças com uma outra crença para gerar um terceiro desejo, e assim por diante. Às vezes, esse processo levará ao desejo de realizar uma ação relativamente simples ou "básica", e esse desejo, por sua vez, fará com que o agente execute a ação básica sem a intervenção de outros desejos mais. Desejos produzidos ou sustentados por esse processo de raciocínio prático são desejos *instrumentais* – a pessoa os tem por que pensa que satisfazê-los levará a outra coisa que ela deseja. Mas nem **todos** os desejos podem ser instrumentais. Se é preciso evitar a circularidade ou uma regressão infinita, então deve haver certos desejos que **não** são realizados porque o agente pensa que satisfazê-los facilitará a satisfação de outro desejo. Tais desejos que não são produzidos ou sustentados pelo raciocínio prático são os desejos **últimos** do agente – pois os objetos dos desejos últimos, o estado de coisas desejado, fazem-se desejados por si mesmos. Um comportamento é **motivado** por um desejo último específico quando esse desejo fizer parte do processo de raciocínio prático que conduz a esse comportamento.

Se por vezes as pessoas têm desejos últimos pelo bem-estar dos outros, e esses desejos motivam o seu comportamento, então o altruísmo é a visão correta, e o egoísmo é falso. No entanto, se *todos* os desejos últimos são de interesse próprio, então o egoísmo é a visão correta e o altruísmo é falso. O esforço para estabelecer uma ou outra dessas opções deu origem a uma literatura empírica vasta e enormemente

---

[27] Para uma demonstração clássica dessa consideração acerca do raciocínio prático, consulte Goldman (1970).

sofisticada. Para uma visão geral dessa literatura, veja o verbete “as abordagens empíricas do altruísmo”.

6. Discordância moral

Dado que o desacordo moral – sobre aborto, digamos, ou sobre pena capital – se mostra, não raro, intratável, haveria alguma razão para supor que problemas morais admitem resoluções objetivas? Embora essa dificuldade seja moeda antiga, a discussão filosófica contemporânea viu-se estimulada pelo “argumento da relatividade” de Mackie (1977, p.36-8) ou, como é chamado por autores posteriores, pelo “argumento da discordância” (BRINK, 1989, p.197; LOEB, 1998). Sobre as diferenças “radicais” no juízo moral, que são comumente observadas, Mackie (1977, p. 36) demonstra “tornarem difícil tratar tais julgamentos como apreensões de verdades objetivas”.

Mackie supôs que seu argumento debilita o *realismo moral*, ou seja, a visão de que, como Smith (1994, p. 9; 13) pontua,

> as questões morais têm respostas corretas, que as respostas corretas são corrigidas por fatos morais objetivos [...] e [...] ao nos engajarmos em argumentos morais, podemos descobrir quais são esses fatos morais objetivos[28] .

Essa noção de objetividade, como Smith reconhece, requer *convergência* entre as visões morais – espera-se que o tipo certo de argumento, de reflexão e de discussão resulte em um acordo moral muito substancial (SMITH, 1994, p. 6)[29] .

---

[28] Tal como acontece com outros ‘ismos’ filosóficos, o caprichoso apelido de “realismo moral” cobre uma substancial diversidade de opiniões. No entanto, muitos realistas morais aceitariam algo próximo à caracterização de objetividade moral de Smith (vide Railton, 1986a, 1986b; Sayre-McCord, 1988a, 2015; Boyd, 1988; Sturgeon, 1988; Brink, 1989; Enoch, 2011; Cuneo, 2014, cap. 1).

[29] Recentemente, alguns realistas morais negaram que o realismo moral esteja comprometido com a convergência. Por exemplo, Shafer-Landau (2003, p. 228), mas confira, ainda, Moody-Adams (1997, p. 109) e Bloomfield (2001), argumenta que “o desacordo não representa nenhuma ameaça ao realismo de qualquer tipo e, portanto, a *fortiori*, não representa nenhuma ameaça ao realismo moral em particular”. Consulte Doris e Plakias (2008) sobre a razão pela qual os realistas não podem se dar ao luxo de ser esperançosos quanto ao desacordo.

Ainda que realistas morais tenham muitas vezes tomado posições bastante otimistas acerca da extensão de um real acordo moral (*vide* Sturgeon, 1988, p. 229; Smith, 1994, p.188), não há como negar a existência de um abundante e persistente desacordo moral; em muitas questões morais há uma notável *falha de convergência*, mesmo depois de uma discussão prolongada. Antirrealistas como Mackie têm uma explicação pronta para esse fenômeno: o juízo moral não é objetivo no sentido de Smith, e não se pode esperar que o argumento moral realize o que Smith e outros realistas pensam que ele possa[30] . Por outro lado, a tarefa do realista é **esclarecer** falhas de convergência: ela deve fornecer uma explicação dos fenômenos que se mostre consistente com o caso em que o juízo moral seja objetivo e o argumento moral seja racionalmente solucionável. Doris e Plakias (2008) a nomeiam como "explicações neutralizadoras". A estratégia do realista visa insistir que o predomínio de um real desacordo moral decorre das limitações dos disputantes ou de suas circunstâncias, persistindo em mostrar que um acordo moral (deveras substancial, se não unânime)[31] **emergiria** em condições **ideais** – aquelas em que, por exemplo, os disputantes fossem totalmente racionais e totalmente informados dos fatos não morais mais relevantes.

Fica, desde logo, evidente que os méritos relativos dessas explicações concorrentes não podem ser determinados com justiça sem considerar de perto os fatores implicados nas reais discordâncias morais. De fato, tal como notaram com acuidade comentadores tanto de simpatias realistas (Sturgeon, 1988, p. 230) quanto antirrealistas (Loeb, 1998, p. 284), o argumento da discordância não pode ser avaliado apenas com recursos filosóficos *a priori*; faz-se preciso, como Loeb observa,

---

[30] *Vide* Williams (1985, p. 136): "numa investigação científica, deveria haver idealmente certa convergência acerca de uma resposta, em que a melhor explicação da convergência envolveria a ideia de que tal resposta representa o modo de ser das coisas; na área da ética, ao menos em um alto nível de generalidade, não há tal esperança de coerência".

[31] É plausível supor que a convergência não requer unanimidade total. Contudo, tal qualificação plausível levanta questões difíceis: quanta dissidência pode ser verificada em condições ideais até que "discordância substancial" seja uma caracterização mais adequada do que "convergência-nem-toda-unânime"? Como é usual em filosofia, não se esperar muita precisão sobre porcentagens, mas é possível suspeitar que a noção relevante de convergência – sempre lembrando que isso ocorre em condições ideais – deve ser pensada para permitir apenas uma mínima dissidência.

"uma grande quantidade de pesquisas empíricas adicionais sobre as circunstâncias e crenças de várias culturas". Tal pesquisa é exigida não só para estimar com precisão a extensão da discordância real, mas também para determinar **por que** o desacordo persiste ou se dissolve. Só então as tentativas dos realistas para "esclarecerem" a discordância moral podem ser avaliadas de maneira justa.

Richard Brandt, que foi um pioneiro no esforço de integrar a teoria ética e as ciências sociais, observou, de início, como a antropologia ajuda a determinar de que maneira as atitudes morais podem convergir sob circunstâncias idealizadas. Obviamente, é bem conhecida a abrangência, na antropologia, de um substancial corpo de trabalho, tais como o dos estudos clássicos de Westermarck (1906) e Sumner (1934), que detalham as perspectivas morais radicalmente divergentes encontradas em culturas ao redor do mundo. Mas, como reconheceu Brandt (1959, p. 283-284), as etnografias típicas acabam por não apoiar inferências confiáveis acerca da convergência de atitudes sob condições ideais, em grande parte porque muitas vezes essas etnografias oferecem orientações limitadas acerca de quanto da discordância moral pode ser atribuído ao desacordo sobre questões factuais que não são de natureza moral, como aquelas relativas a visões religiosas ou cosmológicas.

Com esse tipo de dificuldade em mente, Brandt (1954) empreendeu seu próprio estudo antropológico do povo Hopi, no sudoeste americano, encontrando questões para as quais parecia haver uma séria discordância moral entre atitudes típicas do povo Hopi e do branco americano, que não podiam ser atribuídas de maneira plausível a diferenças de crença sobre fatos não morais[32] . Um exemplo notável é a atitude Hopi em relação ao sofrimento animal, uma atitude que pode chegar a perturbar muitos não-Hopis:

> [As crianças Hopi] às vezes capturam pássaros e fazem deles "animais de estimação". Eles podem ser amarrados a uma corda, para serem forçados a certas "brincadeiras". É um jogo difícil, e os pássaros

[32] Brandt não é o único filósofo que trabalha na tradição "analítica" anglo-americana a produzir etnografia. Ladd (1957) relata seu trabalho de campo com os navajos; suas conclusões (*vide* 1957, p. 328) sobre as dificuldades colocadas pela discordância parecem algo mais otimistas do que as de Brandt, embora talvez não radicalmente em desacordo entre si.

> raramente sobrevivem por muito tempo. [Segundo um informante:] "Às vezes eles cansam e morrem. E ninguém se contrapõe a isso". (Brandt, 1954, p.213)

Brandt (1959, p. 103) fez um esforço conjunto para determinar se essa diferença na perspectiva moral poderia ser atribuída à discordância sobre fatos não morais, mas não conseguiu encontrar uma explicação plausível desse tipo; seus informantes Hopi, por exemplo, não acreditavam que os animais fossem incapazes de sentir dor, nem possuíam crenças cosmológicas que explicassem a aparente crueldade da prática, crenças tais como a de que os animais seriam recompensados pelo martírio na vida após a morte. A melhor explicação para os julgamentos morais divergentes, concluiu Brandt (1954, p. 245; 284), é uma "diferença básica de atitude", já que "grupos às vezes fazem avaliações divergentes quando têm crenças idênticas sobre os objetos".

O argumento de Moody-Adams destaca que, do trabalho etnográfico de Brandt – e de tantos outros –, obtém-se pouca implicação filosófica. Ao desenvolver a doutrina de "significado situacional", da psicologia da Gestalt (por exemplo, DUNKER, 1939), Moody-Adams (1997, p. 34-43) pôde afirmar que todas as instituições, enunciados e comportamentos têm significados que são peculiares ao seu meio cultural, de modo que não podemos estar seguros de que os participantes, em discordâncias transculturais, estejam falando sobre a mesma coisa[33] . O problema do significado situacional, segundo ela, prenuncia certa dificuldade metodológica "insuperável" aos que afirmam existir uma intratável discordância intercultural (1997, p. 36). Os defensores de projetos etnográficos provavelmente responderão, não sem razão, pensamos, que a observação e a entrevista judiciosas, como as que Brandt almejou, **podem** motivar avaliações confiáveis da diversidade avaliativa. Suponhamos, contudo, que Moody-Adams esteja certa, e que tais dificuldades metodológicas sejam intransponíveis. Haveria, então, uma distribuição equitativa da dificuldade: se a observação e a entrevista são realmente tão problemáticas quanto Moody-Adams sugere, **nem** a opinião dos realistas **nem** a dos antirrealistas,

---

[33] Para observações sobre o significado situacional em afinidade com o que se segue, ver Snare (1980, p. 356-362). Agradeço a Walter Sinnott-Armstrong por nos indicar a valiosa discussão de Snare.

sobre a discordância, poderia ser apoiada pelo recurso à evidência empírica. Não pensamos que esse impasse ocorra, porque achamos excessivo o pessimismo metodológico implicado. Cremos que o trabalho empírico sério pode nos dizer muito sobre as culturas e as diferenças entre elas. A maneira apropriada de proceder está em dedicar muita atenção a estudos específicos, ao que eles mostram e deixam de mostrar[34] .

Como Brandt (1959, p. 101-102) reconheceu, a literatura antropológica de sua época nem sempre fornecia tanta informação, acerca dos contornos e das origens exatas de atitudes e crenças morais, quanto os filósofos, que se perguntavam sobre as perspectivas de convergência, gostariam. No entanto, a psicologia social e a ciência cognitiva têm produzido, recentemente, pesquisas que prometem uma discussão aprofundada; durante os últimos 35 anos, houve uma explosão de "psicologia cultural" investigando os processos cognitivos e emocionais de diferentes culturas (Shweder; Bourne, 1982; Markus; Kitayama, 1991; Ellsworth, 1994; Nisbett; Cohen, 1996; Nisbett, 1998, 2003; Kitayama; Markus, 1999; Heine, 2008; Kitayama, Cohen, 2010; Henrich, 2015). Aqui, enfocaremos algumas diferenças culturais encontradas perto de (nossa) casa – diferenças descobertas por Nisbett e seus colegas enquanto examinavam padrões regionais de violência no Norte e no Sul dos Estados Unidos. Queremos demonstrar que esses achados apoiam as conclusões pessimistas de Brandt sobre a probabilidade de convergência no julgamento moral.

A pesquisa do grupo Nisbett pode ser vista como uma aplicação das ferramentas da psicologia social cognitiva à "cultura da honra", um fenômeno que os antropólogos documentaram em várias comunidades ao redor do mundo. Embora essas comunidades difiram em muitos aspectos, elas manifestariam importantes pontos em comum:

> Um aspecto fundamental da cultura da honra é a importância dada ao insulto e à necessidade de responder a ele. Um insulto implica que o alvo é fraco o suficiente para ser intimidado. Como a reputação de força é essencial na cultura da honra, o indivíduo que insulta alguém deve ser forçado a se retratar; se

[34] Para uma discussão mais detalhada sobre a visão de Moody Adams, consulte Doris e Plakias (2008).

> o instigador se recusar, ele deve ser punido — com violência ou mesmo com a morte. (NISBETT; COHEN, 1996, p. 5)

De acordo com Nisbett e Cohen (1996, p. 5-9), um fator importante para a gênese da cultura de honra do Sul encontra-se na presença de uma economia de pastoreio. As culturas de honra são particularmente propensas a se desenvolverem onde os recursos são passíveis de roubo e onde o aparato coercitivo do Estado não pode ser utilizado para prevenir ou punir o roubo. Estas condições ocorrem frequentemente em áreas relativamente remotas, onde o pastoreio é a principal forma de subsistência; a "portabilidade" dos animais de rebanho os torna propensos ao roubo. Em áreas onde predomina a agricultura em vez do pastoreio, a cooperação entre vizinhos é mais importante, infraestruturas governamentais fortes são mais comuns, e os recursos – como as, evidentemente, não transportáveis terras agrícolas – tornam-se mais difíceis de roubar. Em tais economias sociais agrárias, as culturas de honra tendem a não se desenvolver. O sul americano foi originalmente colonizado sobretudo pelos povos de áreas remotas da Grã-Bretanha. Como suas terras natais eram geralmente impróprias para a agricultura, esse povo se constituí, historicamente, por pastores; ao emigrarem da Grã-Bretanha para o sul americano, eles procuraram inicialmente regiões remotas adequadas para pastoreio e, nessas regiões, a cultura da honra floresceu.

No Sul contemporâneo, a polícia e outros serviços governamentais encontram-se amplamente disponíveis e o pastoreio praticamente desapareceu como um modo de vida, o que não impede que certos tipos de violência continuem a ser mais comuns do que no Norte. Nisbett e Cohen (1996) sustentam que os padrões de violência no Sul, assim como as atitudes em relação à violência, insultos e afrontas à honra, são melhor explicados pela hipótese de que uma cultura de honra persiste entre atuais sulistas brancos não-hispânicos. Para apoiar esta hipótese, os autores oferecem um conjunto convincente de evidências, que incluem:

- dados demográficos que indicam que, (1) entre os brancos do sul, as taxas de homicídio são mais altas em regiões mais suscetíveis ao pastoreio do que à agricultura, e que (2) os homens brancos do sul são muito mais propensos, do que homens brancos de outras regiões, a se envolverem em homicídios resultantes de argumentos, embora eles

não sejam mais propensos a se envolverem em homicídios que ocorrem durante um roubo ou outro crime (Nisbett, Cohen, 1996, Cap.2)

- dados de pesquisas que indicam que os sulistas brancos são mais propensos do que os nortistas a acreditarem que a violência seria "extremamente justificada" como resposta a uma variedade de insultos, e se um homem não responde violentamente a tais insultos, ele "não seria muito homem". (Nisbett, Cohen, 1996, Cap.3)

- pesquisas jurídicas indicando que os estados do sul "dão aos cidadãos mais liberdade para usarem da violência na defesa de si mesmos, suas casas e suas propriedades" do que os estados do norte (Nisbett, Cohen, 1996, Cap.5, p.63)

Dois estudos experimentais — um de campo, outro em laboratório — são especialmente impressionantes.

No caso do estudo de campo (Nisbett, Cohen, 1996, p. 73-5), cartas de consulta foram enviadas a centenas de empregadores nos Estados Unidos. As cartas eram supostamente de um homem trabalhador de 27 anos, natural de Michigan, que apresentava uma única mancha em seu sólido histórico. Numa das versões, o "requerente" revelou que havia sido condenado por homicídio culposo. O requerente explicava que brigou com um homem que o confrontara em um bar e dissera aos espectadores que "ele e minha noiva estavam dormindo juntos. Ele riu de mim na minha cara e me pediu para sair se eu fosse homem o suficiente". Segundo o que dizia a carta, o inimigo do requerente foi morto na briga que se seguiu. Numa outra versão, o requerente revelava ter sido condenado por furto de veículo motorizado, praticado devido à necessidade de dinheiro para a família. Nisbett e seus colegas avaliaram 112 cartas de resposta e descobriram que os empregadores do Sul se mostravam significativamente mais propensos a, de forma complacente, cooperarem em resposta à carta do homicida do que os empregadores do Norte, ao passo que nenhuma diferença regional foi encontrada nas respostas à carta de roubo. Um empregador do Sul respondeu à carta do homicida da seguinte forma:

> Quanto aos seus problemas do passado, qualquer pessoa provavelmente poderia estar na situação em que você esteve. Foi apenas um incidente infeliz que não deveria ser usado contra você. Sua honestidade mostra que você é sincero… Desejo-lhe sorte para o seu futuro. Você tem uma atitude positiva e vontade de trabalhar. Essas são qualidades que as empresas procuram nos funcionários. Assim que estiver instalado, se estiver por perto, por favor, não deixe de nos visitar. (Nisbett, Cohen, 1996, p.75)

Nenhuma carta dos empregadores do Norte se revela simpática a esse ponto com o homicida.

No caso do estudo em laboratório (Nisbett, Cohen, 1996, p. 45-48), os participantes, homens brancos dos estados do Norte e do Sul que frequentavam a Universidade de Michigan, foram informados de que amostras de saliva seriam coletadas para medir o açúcar no sangue enquanto realizavam várias tarefas. Logo após a coleta inicial, um participante inocente caminhava por um corredor estreito onde um cúmplice experimental fingia trabalhar em algum arquivo. O cúmplice esbarrou no participante que, fingindo aborrecimento, o chamou de "babaca". Minutos depois, amostras de saliva foram coletadas para determinar o nível de cortisol (um hormônio associado a altos níveis de estresse, ansiedade e excitação) e de testosterona (associado à agressão e ao comportamento de dominância). Como indica a Figura 1, os indivíduos do Sul revelaram aumentos dramáticos nos níveis de cortisol e testosterona, enquanto os do Norte exibiram mudanças muito menores.

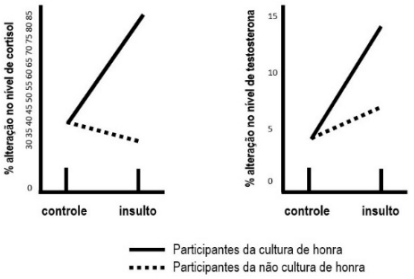


Figura 1.

Os dois estudos assim relatados sugerem que os sulistas respondem mais fortemente a insultos do que os nortistas, tendo com isso uma visão mais simpática de outros que o fazem, ao manifestarem a forma de agir que supostamente tipifica as culturas de honra. Acreditamos que os dados reunidos por Nisbett e seus colegas tornam-se uma amostra persuasiva de que a cultura de honra persiste no sul dos Estados Unidos. Aparentemente, essa cultura afeta os julgamentos, as atitudes e emoções, bem como os comportamentos e até mesmo as respostas fisiológicas das pessoas. Além disso, há evidências de que as práticas de criação dos filhos desempenham papel significativo na transmissão dessa cultura da honra de uma geração para a outra, e que as leis relativamente permissivas acerca de posse de armas, autodefesa e punição corporal nas escolas refletem e reforçam essa cultura de honra do Sul (Nisbett, Cohen, 1996, p.60-63, p.67-9). Em suma, a cultura da honra parece estar profundamente enraizada na cultura sulista contemporânea, apesar de muitas daquelas condições materiais e econômicas que a originaram já não serem encontradas mais tão amplamente[35] .

Cremos que as diferenças culturais Norte/Sul, levantadas por Nisbett e seus colegas, apoiam a conclusão de Brandt de que as atitudes morais raramente convergem, mesmo sob condições ideais. Esses dados deveriam ser especialmente preocupantes para um realista, já que apesar das diferenças que temos relatado, supõem-se haver muito mais em comum entre norte-americanos e sul-americanos contemporâneos – quanto às circunstâncias de linguagem, crença e ideologia – do que, digamos, os ianomâmis e os parisienses. Desse modo, se há pouca razão para esperar convergência nestes casos, provavelmente haverá também em muitos outros.

Fraser e Hauser (2010) não estão convencidos por nossa interpretação dos dados de Nisbett e Cohen. Eles sustentam que, embora esses dados indiquem que nortistas e sulistas diferem na força de sua desaprovação à violência provocada por insultos, eles não mostram que nortistas e sulistas tenham um desacordo moral

---

[35] A última cláusula é importante, já que os realistas (por exemplo, BRINK, 1989, p.200) argumentam às vezes que a aparente discordância moral pode resultar de culturas que aplicam valores morais semelhantes a diferentes condições econômicas (por exemplo, diferenças de atitudes em relação aos doentes e idosos entre culturas pobres e ricas). Mas essa explicação parece ter pertinência duvidosa para as diferenças descritas entre nortistas e sulistas *contemporâneos*, por serem estas adequadamente interpretadas como requerendo *valores diferentes* para *condições econômicas semelhantes*.

real. Desenvolvem ainda a argumentação de que o trabalho de Abarbanell e Hauser (2010) fornece um exemplo muito mais persuasivo de uma discordância moral sistemática entre pessoas de diferentes grupos culturais. Abarbanell e Hauser se concentraram nos juízos morais produzidos pelo povo Maia da zona rural do estado mexicano de Chiapas. Com isso, descobriram que as pessoas naquela comunidade não julgam certas *ações* danosas como sendo piores do que certas *omissões* (falhas em agir) que causam danos idênticos, ao passo que os Maias de zonas urbanas próximas e os usuários ocidentais da Internet julgam ações substancialmente piores do que omissões.

Ainda que a interpretação dos dados de Nisbett e Cohen feita por Fraser e Hauser não nos pareça convincente, concordamos que o estudo de Abarbanell e Hauser providencia um exemplo atraente de certa diferença cultural sistemática em termos de julgamento moral. Barrett et al. (2016) fornece outro exemplo. Seu estudo observou até que ponto a intenção do agente afeta os juízos morais das pessoas em oito sociedades tradicionais, de menor escala, e duas sociedades ocidentais, uma urbana e outra rural. Eles descobriram que em algumas dessas sociedades, notadamente incluindo ambos os grupos ocidentais, a intenção do agente teve um efeito importante, enquanto em outras sociedades a intenção do agente teve pouco ou nenhum efeito.

Como destacamos desde o início, os realistas, que defendem conjecturas sobre a convergência moral, podem se esforçar por *explicar* a diversidade avaliativa com argumentos que a indiquem como devendo ser atribuída às deficiências dos debatedores ou de suas circunstâncias. Se tal estratégia puder ser bem-sucedida, o realismo moral poderá sobreviver a um argumento que esclareça empiricamente o desacordo: tanto pior para a circunstância da reflexão e da discussão moral em questão, menos pior para a objetividade da moralidade. Visto não podermos examinar aqui todas as nuances dessa sugestão, podemos brevemente observar algumas das suas formas mais comuns. Para evitar sermos vagos, concentraremos nossa observação no estudo de Nisbett e Cohen.

Imparcialidade. A estratégia preferida dos realistas morais, interessados em explicar a discordância moral, está em dizer que este desacordo deriva dos efeitos distorcidos do interesse individual (ver STURGEON, 1988, p.229-230; ENOCH, 2009, p.24-29); talvez a persistência do desacordo não denuncie tanto os traços profundos do argumento e do julgamento moral como faz a obstinação com que os indivíduos buscam obter vantagem. Por exemplo, disputas aparentemente morais sobre a distribuição de riqueza podem derivar de percepções – talvez em

grande parte rudimentares – com interesses individuais e de classe, antes que de divergências quanto aos princípios da justiça. Em tais circunstâncias, essa discordância moral persistente não cumpre a condição de imparcialidade e, por isso, não incomoda o realista moral. Contudo, é um tanto implausível sugerir que as discordâncias Norte/Sul, sobre em que momento a violência é justificada, não cumprem a condição de imparcialidade. Não parece haver razão para se pensar que os sulistas estariam relutantes em universalizar seus julgamentos sobre indivíduos relativamente semelhantes em circunstâncias relativamente semelhantes, como de fato sugere o "estudo das cartas" de Nisbett e Cohen. É possível defender um código de honra violento sem recorrer a alguma alegação especial[36] . Não pretendemos, com isso, difamar os valores do Sul; nosso ponto é que, embora possa haver boas razões para criticar a cultura da honra sulista, não é óbvio que a razão possa ser a falta de imparcialidade, se a imparcialidade for (grosseiramente) entendida enquanto alinhada a uma vontade de universalizar os próprios julgamentos morais.

Consciência plena e vívida de fatos não morais relevantes. Realistas morais argumentam que as discordâncias morais derivam muitas vezes do desacordo sobre questões não morais. De acordo com Boyd (1988, p. 213; *vide* Brink, 1989, p. 202-203; Sturgeon, 1988, p. 229),

> Um cuidadoso exame filosófico revelará [...] que o acordo sobre questões não morais eliminaria *quase todo* desacordo sobre os tipos de questões morais que surgem na prática moral comum.

Esta é uma conjectura plausível para os dados que acabamos de considerar? Achamos difícil imaginar que o acordo sobre fatos não morais poderia servir de artifício, já que podemos facilmente imaginar nortistas e sulistas estando em pleno acordo sobre fatos não morais relevantes nos casos descritos. Membros dos dois grupos provavelmente concordariam que o candidato a emprego foi traído, por

---

[36] Os estudos jurídicos, que Nisbett e Cohen (1996, p. 57-78) revisaram, deixam claro que as legislaturas do Sul quase sempre estão dispostas a promulgar leis que reflitam a visão da cultura da honra em relação às circunstâncias sob as quais a violência é justificada, de modo a sugerir que há pelo menos algum apoio entre os sulistas para a ideia de que os valores de honra devam ser universalizáveis.

exemplo, ou que chamar alguém de "idiota" é um insulto. Cremos ser bem mais plausível supor que a discordância resida em atitudes avaliativas diferentes e profundamente arraigadas em relação às respostas apropriadas quanto à traição, ao desafio e ao insulto.

Leitores filosóficos experientes observarão rapidamente que termos como "desafio" e "insulto" parecem termos éticos "grosseiros", em que o avaliativo e o descritivo se misturam (*vide* WILLIAMS, 1985, p. 128-130); em razão disso, é muito difícil dizer qual seria a extensão da discordância factual. Na verdade, isso é de pouca ajuda para o recurso que estamos considerando, uma vez que a resposta do desacordo-sobre-fatos-não-morais aparentemente **exige** que seja **possível** desvincular discordância factual e moral.

É claramente possível que a consciência plena e vívida dos fatos não morais possa motivar o tipo de mudança esperada nas atitudes sulistas pelo realista moral (ao menos o do Norte). Se os sulistas se tornassem vivamente conscientes de que sua cultura de honra estava implicada em sua violência, eles poderiam ser instigados a mudarem sua perspectiva moral. (Consideramos esta maneira de colocar o exemplo como a mais natural, mas nada filosófico estaria implicado aqui. Se preferir, substitua a possibilidade de os nortistas endossarem valores de honra após uma exposição aos fatos.) Por outro lado, os sulistas poderiam insistir que os valores da honra devem ser alimentados mesmo sendo seu custo a promoção da violência; o lema "morte em vez de desonra", afinal, tem uma história longa e honrosa. Para nós, o ônus da argumentação recai sobre o realista que afirma — **a despeito da cultura e da história** — que os sulistas chegariam a mudar de ideia se estivessem plenamente cientes dos fatos pertinentes.

Libertação da "Anormalidade". Realistas poderiam argumentar que um bom número de discordâncias morais resulta de falhas de racionalidade da parte dos debatedores (BRINK, 1989, p. 199-200). Obviamente, o desacordo decorrente de deficiências cognitivas não se torna embaraçoso para o realismo moral; em suma, que certo desacordo persista quando algumas ou todas as partes em disputa revelam-se bastante insanas nada mostra de profundo sobre moralidade. Contudo, não parece plausível que atitudes sulistas mais lenientes em relação a certas formas de violência sejam prontamente atribuídas a uma incapacidade cognitiva generalizada. Por certo, trata-se de uma questão empírica, mas não conhecemos nenhuma evidência sugerindo que sulistas sofram de alguma deficiência cognitiva que os impeça de entenderem os fatores demográficos e comportamentais na gênese da violência, ou de entenderem qualquer outro assunto sobre fatos. O que é necessário

para levantar uma acusação de irracionalidade seria a evidência de comprometimento cognitivo **independente** das diferenças de atitude, e mais evidências de que esse comprometimento está implicado na adesão aos valores em disputa. Neste caso, como em tantos outros, temos dificuldade em observar como as acusações de anormalidade ou irracionalidade podem ser feitas sem que um lado exija a questão contra o outro.

O trabalho de Nisbett e seus colegas pode representar um contraexemplo poderoso a qualquer teoria que sustente que o argumento racional tende a convergir em questões morais importantes; a evidência sugere que as diferenças entre as atitudes Norte/Sul em relação à violência e à honra podem persistir mesmo sob o tipo de condições ideais em consideração. Admitimos que tais conclusões devam ser provisórias. No aspecto filosófico, nem todas as estratégias plausíveis para "explicar" a discordância moral e fundamentar as expectativas de convergência foram consideradas[37] . No aspecto empírico, este verbete relatou apenas alguns estudos, e aqueles que foram considerados, como qualquer trabalho empírico, podem ser criticados por motivos conceituais ou metodológicos[38] . Finalmente, deve ficar claro o que este verbete *não* está afirmando: quaisquer conclusões aqui – mesmo se adequadamente obtidas – não são uma "refutação" de todas as versões de realismo moral, já que existem versões de realismo moral que não requerem convergência (BLOOMFIELD, 2001; SHAFER-LANDAU, 2003). Em vez disso, essa discussão deveria dar uma ideia do trabalho empírico que os filósofos têm pela frente, se quiserem fazer conjecturas defensáveis acerca da discordância moral.

---

[37] Além dos expedientes considerados aqui, os realistas podem apelar com razão para, **entre outras coisas**, requisitos de coerência interna e diferentes "níveis" de pensamento moral (teórico X popular, abstrato X concreto, geral X particular) nos quais a discordância moral pode ou não se manifestar. Brink (1989, p. 197-210) e Loeb (1998) oferecem discussões valiosas com muito mais detalhes do que pudemos dispor aqui: Brink manifestando simpatias realistas e Loeb tendendo ao antirrealismo.

[38] Nisbett e Cohen provavelmente se sairão muito bem sob tal escrutínio. Ver a leitura favorável de Tetlock (1999).

7. Conclusão

O progresso, na teorização ética, quase sempre requer o progresso em questões psicológicas difíceis sobre o que se pode esperar que os seres humanos realizem em contextos morais. Não é surpresa, então, que a psicologia moral seja uma área central de investigação na ética filosófica. Também não deveria ser surpreendente que pesquisas empíricas, como aquelas conduzidas nos departamentos de psicologia, possam auxiliar, substancialmente, tal investigação. Tampouco deve surpreender que a pesquisa em psicologia moral tenha se tornado **metodologicamente pluralista**, explorando recursos de várias disciplinas e procurando contribuir para elas. Ilustramos, aqui, como essa investigação interdisciplinar pode prosseguir em relação àqueles problemas centrais da ética filosófica.

,

Referência bibliográfica

ABARBANELL, L., Hauser, M. D. “Mayan Morality: An Exploration of Permissible Harms”, *in Cognition*, 115(2), p.207–224, 2010.

ADAMS, R. M. **A Theory of Virtue**: Excellence in Being for the Good. Oxford, Oxford University Press, 2006.

ALFANO, M. **Character as Moral Fiction**, Cambridge, Cambridge University Press, 2013.

ALFANO, M. **Moral Psychology**: An Introduction. Cambridge, Polity Press, 2016.

ANDOW, J., Cova, F. Why Compatibilist Intuitions Are Not Mistaken: A Reply to Feltz and Millan, **Philosophical Psychology**, 29(4), p.550–566. 2016.

ANNAS, J. Comments on John Doris’ *Lack of Character*, in **Philosophy and Phenomenological Research**, 71(3), p.636–642. 2005.

ANNAS, J. **Intelligent Virtue**. Oxford, Oxford University Press, 2011.

ANSCOMBE, G.E.M. “Modern Moral Philosophy”, in *Philosophy*, 33(124), 1–19, 1958.

APPIAH, K. A. **Experiments in Ethics**. Cambridge, MA, Harvard University Press, 2008.

ARISTOTLE. Nichomachean Ethics, in *The Complete Works of Aristotle*, edited by J. Barnes. Princeton, Princeton University Press, 1984.

ARPALY, N. “Comments on *Lack of Character* by John Doris”, in *Philosophy and Phenomenological Research*, 71(3), p. 643–647, 2005.

ATHANASSOULIS, N. “A Response to Harman: Virtue Ethics and Character Traits”, in *Proceedings of the Aristotelian Society*, 100(1), 1999.

BADHWAR, N. K. “The Milgram Experiments, Learned Helplessness, and Character Traits”, in *The Journal of Ethics*, 13(2–3), p. 257–289. 2009.

BARON, J. “Nonconsequentialist Decisions”, in *Behavioral and Brain Sciences*, 17(1), p.1-42, 1994.

BARON, J. **Thinking and Deciding**, 3rd edition, Cambridge: Cambridge University Press, 2001.

BARRETT, H.C. et.al. “Small-Scale Societies Exhibit Fundamental Variation in the Role of Intentions in Moral Judgment”, in *Proceedings of the National Academy of Sciences*, 113(17), p.4688–4693, 2016.

BATSON, C. D. **The Altruism Question**: Toward a Social-Psychological Answer. Hillsdale-NJ, Lawrence Erlbaum Associates, 1991.

BATSON, C. D. **Altruism in Humans**, Oxford, Oxford University Press, 2011.

BEAR, A., KNOBE, J. “What Do People Find Incompatible With Causal Determinism?” in *Cognitive Science*, 40(8), p.2025–2049, 2016.

BESSER-JONES, L. “Social Psychology, Moral Character, and Moral Fallibility”, in *Philosophy and Phenomenological Research*, 76(2), p.310–332, 2008.

BJÖRNSSON, G. “Incompatibilism and ‘Bypassed’ Agency”, in A. R. Mele (ed.), *Surrounding Free Will*, Oxford, Oxford University Press, p. 95–112, 2014.

BJÖRNSSON, G., Pereboom, D. “Traditional and Experimental Approaches to Free Will and Moral Responsibility”, in *Sytsma and Buckwalter*, 2016, p.142–157.

BLOOMFIELD, P. “Virtue Epistemology and the Epistemology of Virtue”, in *Philosophy and Phenomenological Research*, 60(1), p.23-43, 2000.

BLOOMFIELD, P. **Moral Realit.** New York, Oxford University Press, 2001.

BLOOMFIELD, P. “Some Intellectual Aspects of the Cardinal Virtues”, in *Oxford Studies in Normative Ethics*, volume 3, M. Timmons (ed.), p. 287–313, 2014.

BOORSE, C. “On the Distinction between Disease and Illness”, in *Philosophy and Public Affairs*, 5(1), p.49–68, 1975.

BOYD, R. “How to Be a Moral Realist”, in Sayre-McCord 1988b: 181–228, 1988.

BRANDT, R. B. **Hopi Ethics**: A Theoretical Analysis. Chicago, The University of Chicago Press, 1954.

BRANDT, R. B. **Ethical Theory**: The Problems of Normative and Critical Ethics. Englewood Cliff-NJ, Prentice-Hall, 1959.

BRATMAN, M. E. “Identification, Decision, and Treating as a Reason”, in *Philosophical Topics*, 24(2), p.1-18, 1996.

BRINK, D. O. **Moral Realism and the Foundations of Ethics**. Cambridge, Cambridge University Press, 1989.

BROAD, C.D. **Five Types of Ethical Theory.** New York, Harcourt, Brace, 1930.

BROAD, C.D. "Egoism as a Theory of Human Motives", in *The Hibbert Journal*, 48, p.105–114, 1950 [Republicado em seu *Ethics and the History of Philosophy: Selected Essays.* London, Routledge, KEGAN, P., 1952, p. 218–231]

CAMERON, C. D., Payne B. K., Doris, J. "Morality in High Definition: Emotion Differentiation Calibrates the Influence of Incidental Disgust on Moral Judgments", in *Journal of Experimental Social Psychology*, 49(4), p.719-725, 2013.

CAMPBELL, C.A. "Is 'Freewill' a Pseudo-problem?", in *Mind*, 60(240), p.441-465, 1951.

CAPPELEN, H. **Philosophy Without Intuitions**. Oxford, Oxford University Press. 2012.

CERVONE, D., Shoda, Y (eds.). **The Coherence of Personality**: Social-Cognitive Bases of Consistency, Variability, and Organization. New York and London, Guilford Press, 1999.

CHIESI, H. L., Spilich, G. J., Voss, J. F. "Acquisition of domain-related information in relation to high and low domain knowledge", in *Journal of verbal learning and verbal behavior*, 18(3), p.257–273, 1979.

CIALDINI, R. B., et. al. "Reinterpreting the Empathy-Altruism Relationship: When One into One Equals Oneness", in *Journal of Personality and Social Psychology*, 73(3), p. 481-494, 1997.

COVA, F., KITANO, Y. "Experimental Philosophy and the Compatibility of Free Will and Determinism: A Survey", in *Annals of the Japan Association for Philosophy of Science*, 22, p.17–37, 2013.

COVA, F. et. al. "Judgments about Moral Responsibility and Determinism in Patients with Behavioural Variant of Frontotemporal Dementia: Still Compatibilists", in *Consciousness and Cognition*, 21(2), p.851–864, 2012.

CUNEO, T. **Speech and Morality: On the Metaethical Implications of Speaking**. Oxford, Oxford University Press, 2014.

DARLEY, J. M., Batson, C. D. "'From Jerusalem to Jericho': A Study of Situational and Dispositional Variables In Helping Behavior", in *Journal of Personality and Social Psychology*, 27(1), p.100–108, 1973.

DECETY, J., Wheatley, T. (eds.) **The Moral Brain**: A Multidisciplinary Perspective. Cambridge-MA, MIT Press, 2015.

DEERY, O. Nahmias, E. "Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood", in *Philosophical Studies*, 174(5), p.1255–1276, 2017.

DENNETT, D. C. **Elbow Room**: The Varieties of Free Will Worth Wanting. Cambridge, MA: MIT Press, 1984.

DEPAUL, M. "Character Traits, Virtues, and Vices: Are There None?" in *Proceedings of the 20th World Congress of Philosophy, v. 1*, Bowling Green, OH: Philosophy Documentation Center, p. 141–157, 1999.

DEUTSCH, M. **The Myth of the Intuitive**: Experimental Philosophy and Philosophical Method. Cambridge, MA: MIT Press, 2015.

DIXON, T. **The Invention of Altruism**: Making Moral Meanings in Victorian Britain. Oxford: Oxford University Press, 2008.

DONNELLAN, M. B., LUCAS, R.E., FLEESON, W. (eds.) "Personality and Assessment at Age 40: Reflections on the Past Person-Situation Debate and Emerging Directions of Future Person-Situation Integration and Assessment at Age 40", in *Journal of Research in Personality*, special issue, 43(2), p.117–290, 2009.

DORIS, J. M. "Persons, Situations, and Virtue Ethics", in *Noûs*, 32(4), p. 504–530, 1998.

DORIS, J. M. **Lack of Character**: Personality and Moral Behavior. New York, Cambridge University Press, 2002.

DORIS, J. M. "Précis" and "Replies: Evidence and Sensibility", in *Philosophy and Phenomenological Research*, 71(3), p.632–5, p. 656–77, 2005.

DORIS, J. M. "Out of Character: On the Psychology of Excuses in the Criminal Law", in H. LaFollette (ed.), *Ethics in Practice*, third edition, Oxford: Blackwell Publishing, 2006.

DORIS, J. M. "Heated Agreement: Lack of Character as Being for the Good", in *Philosophical Studies*, 148(1), p.135-46, 2010.

DORIS, J. M. **Talking to Our Selves**: Reflection, Ignorance, and Agency. Oxford, Oxford University Press, 2015.

DORIS, J. M. **Character Trouble**: Undisciplined Essays on Personality and Agency. Oxford: Oxford University Press [no prelo]

DORIS, J. M. "Making Good: In Search of Moral Expertise" [em preparação].

DORIS, J. M., Plakias, A. "How to Argue about Disagreement: Evaluative Diversity and Moral Realism", in Sinnott-Armstrong (2008b), p. 303–353, 2008.

DORIS, J. M. Prinz, J.J. "Review of K. Anthony Appiah, *Experiments in Ethics*", in *Notre Dame Philosophical Reviews*, 2009-10-03. <>

DORIS, J. M., The Moral Psychology Research Group (eds.). **The Moral Psychology Handbook.** Oxford, Oxford University Press, 2010.

DORIS, J. M., Stich, S P. "As a Matter of Fact: Empirical Perspectives on Ethics", in Frank Jackson and Michael Smith (eds.), *The Oxford Handbook of*

Contemporary Philosophy, Oxford: Oxford University Press, 2005.

DUNKER, K. "Ethical Relativity? (An Enquiry into the Psychology of Ethics)", in *Mind*, 48(189), p.39–53, 1939.

ELLSWORTH, Phoebe C., 1994, "Sense, Culture, and Sensibility", in Shinobu Kitayama and Hazel Rose Markus (eds.), *Emotion and Culture: Empirical Studies of Mutual Influence*, Washington: American Psychological Association.

ENOCH, D. "How is Moral Disagreement a Problem for Realism?", in *The Journal of Ethics*, 13(1), p.15–50, 2009.

ENOCH, D. **Taking Morality Seriously**: A Defense of Robust Realism. Oxford, Oxford University Press, 2011.

ERICSSON, K. A. "Why Expert Performance Is Special and Cannot Be Extrapolated From Studies of Performance in the General Population: A Response to Criticisms", in *Intelligence*, 45. p.81-103, 2014.

ERICSSON, k. A. et. al., 1993, "The Role of Deliberate Practice in the Acquisition of Expert Performance", in *Psychological Review*, 100(3), p. 363–406, 1993.

FEINBERG, J. "Psychological Egoism", in *Reason and Responsibility*, Joel Feinberg (ed.), Belmont, CA: Dickenson Publishing. Reprinted in various editions including the tenth, co-edited with Russ Shafer-Landau, Belmont, CA: Wadsworth, 1999. Baseado nos materiais compostos por estudantes da Brown University, 1958.

FELTZ, A., COVA, F. "Moral Responsibility and Free Will: A Meta-Analysis", in *Consciousness and Cognition*, 30, p. 234–246, 2014.

FELTZ, A., MILLAN, M. "An Error Theory for Compatibilist Intuitions", in *Philosophical Psychology*, 28(4), p. 529–555, 2013.

FIGDOR, C., PHELAN, M. "Is Free Will Necessary for Moral Responsibility? A Case for Rethinking Their Relationship and the Design of Experimental Studies in Moral Psychology", in *Mind and Language*, 30(5), p. 603–627. 2015.

FISCHER, J. M. **The Metaphysics of Free Will**. Oxford, Blackwell, 1994.

FLANAGAN, O. **Varieties of Moral Personality**: Ethics and Psychological Realism. Cambridge, MA: Harvard University Press, 1991.

FLANAGAN, O. "Moral Science? Still Metaphysical After All These Years", in Darcia Narvaez and Daniel K. Lapsley (eds.), *Personality, Identity, and Character*, Cambridge: Cambridge University Press, p. 52–78, 2009.

FRANKFURT, H. **The Importance of What We Care About.** Cambridge, Cambridge University Press, 1988.

FRASER, B., HAUSER, M. “The Argument from Disagreement and the Role of Cross-Cultural Empirical Data”, in *Mind and Language*, 25(5), p. 541–560, 2010.

FULFORD, K.W.M. **Moral Theory and Medical Practice**. Cambridge, Cambridge University Press, 1989.

GIGERENZER, G. **Adaptive Thinking**: Rationality in the Real World. New York, Oxford University Press, 2000.

GIGERENZER, G., TODD, P. T., the ABC Research Group. **Simple Heuristics that Make Us Smart**. New York, Oxford University Press, 1999.

GILOVICH, T., GRIFFIN, D.W., KAHNEMAN, D. (eds.) **Heuristics and Biases**: The Psychology of Intuitive Judgment. New York, Cambridge University Press, 2002.

GLICKMAN, Mark E. “A Comprehensive Guide to Chess Ratings”, in *American Chess Journal*, 3, p.59–102, 1995.

GOLDMAN, A. I. **A Theory of Human Action.** Englewood-Cliffs, NJ: Prentice-Hall, 1970.

HAIDT, J., Koller, S. H., Dias, M.G. “Affect, Culture, and Morality, Or Is It Wrong to Eat Your Dog?”, in *Journal of Personality and Social Psychology*, 65(4), p. 613–28, 1993.

HAJI, I. “Compatiblist Views of Freedom and Responsibility”, in Robert Kane (ed.), *The Oxford Handbook of Free Will*, New York, Oxford University Press, 2002.

HAMBRICK, D. Z. et. al. “Deliberate Practice: Is That All It Takes to Become an Expert?”, in *Intelligence*, 45, p. 34–45, 2014.

HARMAN, G. “Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error”, in *Proceedings of the Aristotelian Society*, 99, p. 315–331, 1999.

HARMAN, G. “The Nonexistence of Character Traits”, *Proceedings of the Aristotelian Society*, 100, p.223–226, 2000.

HARMAN, G. “Skepticism about Character Traits”, in *The Journal of Ethics*, 13(2–3), p. 235–242, 2009.

HEINE, S. J. **Cultural Psychology**. New York, W.W. Norton, 2008.

HELZER, E. G., PIZARRO, D. A. “Dirty Liberals! Reminders of Physical Cleanliness Influence Moral and Political Attitudes”, in *Psychological Science*, 22(4), p.517–522, 2011.

HENRICH, J. **The Secret of Our Success**: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter. Princeton, NJ: Princeton University Press, 2015.

HOBBES, T. ***Leviathan****: Edited with an Introduction by C.B. Macpherson*. London, Penguin Books, 1981.

HOROWITZ, T. “Philosophical Intuitions and Psychological Theory”, in Michael R. DePaul and William Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry*, Lanham, Maryland: Rowman and Littlefield, 1998.

HURSTHOUSE, R. **On Virtue Ethics**. Oxford and New York: Oxford University Press, 1999.

ISEN, A. M., LEVIN, P.F. “Effect of Feeling Good on Helping: Cookies and Kindness”, in *Journal of Personality and Social Psychology*, 21(3), p.384–388, 1972.

JACKSON, F. **From Metaphysics to Ethics**: A Defense of Conceptual Analysis. New York, Oxford University Press, 1998.

JACKSON, F. PETTIT, P. “Moral Functionalism and Moral Motivation”, in *Philosophical Quarterly*, 45(178), p.20–40 1995.

JACOBSON, D. “Seeing By Feeling: Virtues, Skills, and Moral Perception”, in *Ethical Theory and Moral Practice*, 8(4), p.387–409, 2005.

JOYCE, R. **The Evolution of Morality**. Cambridge, MA: MIT Press, 2006.

KAHNEMAN, D. **Thinking, Fast and Slow**. New York: Farrar, Straus and Giroux, 2011.

KAHNEMAN, D., SLOVIC, P., TVERSKY, A. **Judgment Under Uncertainty**: Heuristics and Biases. Cambridge: Cambridge University Press, 1982.

KAMTEKAR, R. “Situationism and Virtue Ethics on the Content of Our Character”, in *Ethics*, 114(3), p.458–91, 2004.

KANE, R. **The Significance of Free Will**. Oxford, Oxford University Press, 1996.

KANE, R. “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism”, in *Journal of Philosophy*, 96(5), p. 217–240, 1999.

KANE, R. “Introduction: The Contours of Contemporary Free Will Debates”, in Robert Kane (ed.), *The Oxford Handbook of Free Will*, New York: Oxford University Press, 2002.

KANT, I. *Fundamental Principles of the Metaphysics of Morals*, Translated by Thomas K. Abbott. Englewood Cliffs, NJ: Prentice Hall / Library of Liberal Arts, 1949.

KITAYAMA, S., MARKUS, H.R. “Yin and Yang of the Japanese Self: The Cultural Psychology of Personality Coherence”, in Cervone and Shoda (1999, ch.8), 1999.

KITAYAMA, S., COHEN, D. **Handbook of Cultural Psychology**. New York: Guilford Press, 2010.

KITCHER, P. “Varieties of Altruism”, in *Economics and Philosophy*, 26(2), p.121–148, 2010.

KITCHER, P. 2011, *The Ethical Project*, Cambridge, MA: Harvard University Press.

KNOBE, J. “Intentional Action and Side Effects in Ordinary Language”, in *Analysis*, 63(279), p.190–193, 2003a.

KNOBE, J. “Intentional Action in Folk Psychology: An Experimental Investigation”, in *Philosophical Psychology*, 16(2), p.309–324, 2003b.

KNOBE, J. “The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology”, in *Philosophical Studies*, 130(2), p.203–231, 2006.

KNOBE, J. “Person as Scientist, Person as Moralist”, in *Behavioral and Brain Sciences*, 33(4), p.315–329, 2010.

KNOBE, J. “Free Will and the Scientific Vision”, in Edouard Machery and Elizabeth O’Neill (eds.), *Current Controversies in Experimental Philosophy*, New York and London: Routledge, 2014.

KNOBE, J., LEITER, B. “The Case for Nietzschean Moral Psychology”, in Brian Leiter and Neil Sinhababu (eds.) *Nietzsche and Morality*, Oxford: Oxford University Press. 83–109, 2007.

KRUGER, J., DUNNING, D. “Unskilled and Unaware of It: How Difficulties in Recognizing One’s Own Incompetence Lead to Inflated Self-Assessments”, in *Journal of Personality and Social Psychology*, 77(6), p.1121–1134, 1999.

KUPPERMAN, J. J. “The Indispensability of Character”, in *Philosophy*, 76(02), p.239–50, 2001.

LADD, J. **The Structure of a Moral Code**: A Philosophical Analysis of Ethical Discourse Applied to the Ethics of the Navaho Indians. Cambridge, MA: Harvard University Press, 1957.

LAFOLLETTE, H. (ed.) **The Blackwell Guide to Ethical Theory.** Oxford, Blackwell Publishing, 2000.

LEIKAS, S., LÖNNQVIST, J.-E., VERKASALO, M. “Persons, Situations, and Behaviors: Consistency and Variability of Different Behaviors in Four Interpersonal Situations”, *Journal of Personality and Social Psychology*, 103(6): 1007–1022, 2012.

LERNER, J. S., GOLDBERG, J.H., TETLOCK, F. E. “Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility”, in *Personality and Social Psychology Bulletin*, 24(6), p. 563–574, 1998.

LEWIS, D. “Dispositional Theories of Value”, in *Proceedings of the Aristotelian Society*, 63 (sup), p.113–37, 1989.

LIAO, S. M. et. al. “Putting the Trolley in Order: Experimental Philosophy and the Loop Case”, in *Philosophical Psychology*, 25(5), p.661–671, 2012.

LOEB, Don., 1998, “Moral Realism and the Argument from Disagreement”, *Philosophical Studies*, 90(3): 281–303. doi:10.1023/A:1004267726440

MACHERY, Edouard, 2010, “The Bleak Implications of Moral Psychology”, *Neuroethics*, 3(3): 223–231. doi:10.1007/s12152-010-9063-7

MACHERY, E., DORIS, J.M. “An Open Letter to Our Students: Going Interdisciplinary”, in Voyer and Tarantola [no prelo].

MACINTYRE, A. “Egoism and Altruism”, in Paul Edwards (ed.), *The Encyclopedia of Philosophy*, vol. 2, first edition, New York: Macmillan, p. 462–466, 1967.

MACKIE, J.L., 1977, *Ethics: Inventing Right and Wrong*, New York: Penguin Books.

MACNAMARA, B. N., HAMBRICK, D. Z., OSWALD, F. L. “Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-Analysis”, in *Psychological Science*, 25(8), p.1608–1618, 2014.

MARKUS, H. R., KITAYAMA, S. “Culture and the Self: Implications for Cognition, Emotion, and Motivation”, in *Psychological Review*, 98(2), p.224–253, 1991.

MAY, J. “Psychological Egoism”, *Internet Encyclopedia of Philosophy*, 2011a <>

MAY, J. “Egoism, Empathy, and Self-Other Merging”, in *Southern Journal of Philosophy*, 49(s1), p.25–39, 2011b.

MAY, J. “Relational Desires and Empirical Evidence against Psychological Egoism: On Psychological Egoism”, in *European Journal of Philosophy*, 19(1), p.39–58, 2011c.

MCGRATH, S. “Moral Disagreement and Moral Expertise”, in *Oxford Studies in Metaethics*, volume 3, Russ Shafer-Landau (ed.), New York: Oxford University Press, p. 87–108, 2008.

MCGRATH, S. “Skepticism about Moral Expertise as a Puzzle for Moral Realism”, in *Journal of Philosophy*, 108(3), p.111–137, 2011.

MEHL, M. R. et. al. “Character and Coherence: Testing the Stability of Naturalistically Observed Daily Moral Behavior”, in Miller et al. 2015, p.630–51, 2015.

MELE, A. R. **Free Will and Luck.** New York, Oxford University Press, 2006.

MELE, A. R. “Manipulation, Moral Responsibility, and Bullet Biting”, in *Journal of Ethics*, 17(3), p.167–84, 2013.

MERRITT, M. W. “Virtue Ethics and Stuationist Personality Psychology”, in *Ethical Theory and Moral Practice*, 3(4), p.365–83, 2000.

MERRITT, M. W. “Aristotelean Virtue and the Interpersonal Aspect of Ethical Character”, in *Journal of Moral Philosophy*, 6(1), p.23–49, 2009.

MERITT, M. W., DORIS, J. M., HARMAN, G. “Character”, in Doris et al. 2010, p.355–401, 2010.

MILGRAM, S. **Obedience to Authority.** New York, Harper and Row, 1974.

MILLER, C. B. “Social Psychology and Virtue Ethics”, in *The Journal of Ethics*, 7(4), p.365–92, 2003.

MILLER, C. B. **Moral Character**: An Empirical Theory. Oxford: Oxford University Press, 2013.
MILLER, C. B. **Character and Moral Psychology**. Oxford: Oxford University Press, 2014.
MILLER, C. B. et. al. (eds.) **Character**: New Directions from Philosophy, Psychology, and Theology. New York, Oxford University Press, 2015.
MISCHEL, W. **Personality and Assessment**. New York: John J. Wiley and Sons, 1968.
MISCHEL, W. "Personality Coherence and Dispositions in a Cognitive-Affective Personality System (CAPS) Approach", in Cervone and Shoda (1999, ch. 2), 1999.
MONTMARQUET, J. "Moral Character and Social Science Research", in *Philosophy*, 78(03), p. 355–368, 2003.
MOODY-ADAMS, M. M. **Fieldwork in Familiar Places**: Morality, Culture, and Philosophy. Cambridge, MA: Harvard University Press, 1997.
MURPHY, D. **Psychiatry in the Scientific Image**. Cambridge, MA: MIT Press, 2006.
MURRAY, D., NAHMIAS, E. "Explaining Away Incompatibilist Intuitions", in *Philosophy and Phenomenological Research*, 88(2), p. 434–467, 2014.
MURRAY, D., LOMBROZO, T. "Effects of Manipulation on Attributions of Causation, Free Will, and Moral Responsibility", in *Cognitive Science*, 41(2), p.447–481, 2016.
NADO, J. "The Intuition Deniers", in *Philosophical Studies*, 173(3), p. 781–800, 2016.
NAGEL, T. **The Possibility of Altruism**. Oxford: Oxford University Press, 1970.
NAGEL, T. **The View From Nowhere.** New York and Oxford, Oxford University Press, 1986.
NAHMIAS, E. "Intuitions about Free Will, Determinism, and Bypassing", in Robert Kane (ed.), *The Oxford Handbook of Free Will*, second edition, Oxford: Oxford University Press, 2011.
NAHMIAS, E. et. al. "Is Incompatiblism Intuitive?", in *Philosophy and Phenomenological Research*, 73(1), p. 28–53, 2009.
NICHOLS, S., KNOBE, J. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions", in *Noûs*, 41(4), p. 663–685, 2007.
NISBETT, R. E. "Essence and Accident", in John M. Darley and Joel Cooper (eds.), *Attribution and Social Interaction: The Legacy of Edward E. Jones*, Washington: American Psychological Association, 1998.
NISBETT, R. E. **The Geography of Thought**: How Asians and Westerners Think Differently … and Why. New York: Free Press, 2003.
NISBETT, R. E., Borgida, E. "Attribution and the psychology of prediction", in *Journal of Personality and Social Psychology*, 32(5), p.932–943, 1975.
NISBETT, R. E., Cohen, D. **Culture of Honor**: The Psychology of Violence in the South. Boulder, CO: Westview Press, 1996.

NISBETT, R. E., Ross, L. *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice-Hall, 1980.

O'CONNOR, T. **Persons and Causes**: The Metaphysics of Free Will. New York: Oxford University Press, 2000.

OLIN, L., DORIS, J.M. "Vicious Minds: Virtue Epistemology, Cognition, and Skepticism", in *Philosophical Studies*, 168(3), p. 665–92, 2014.

PEREBOOM, D*. **Living Without Free Will***. Cambridge: Cambridge University Press, 2001.

PEREBOOM, D**. Free Will, Agency, and Meaning in Life**, Oxford: Oxford University Press, 2014.

PETRINOVICH, L., O'NEILL, P. "Influence of Wording and Framing Effects on Moral Intuitions", in *Ethology and Sociobiology*, 17(3), p.145–171, 1996.

PHILLIPS, J., SHAW, A. "Manipulating Morality: Third-Party Intentions Alter Moral Judgments by Changing Causal Reasoning", in *Cognitive Science*, 39(6), p.1320–47, 2014.

PINK, T. **Free Will**: A Very Short Introduction. New York: Oxford University Press, 2004.

PRINZ, J. J. "The Normativity Challenge: Cultural Psychology Provides the Real Threat to Virtue Ethics", in *The Journal of Ethics*, 13(2–3), p. 117–144, 2009.

PUST, J. **Intuitions as Evidence**. New York, Garland Publishing, 2000.

PUST, J. **The Elements of Moral Philosophy**. fourth edition, New York: McGraw-Hill, 2003.

RACHELS, J. "Naturalism", in LaFollette (2000), p.74–91, 2000.

RAILTON, P. "Facts and Values", in *Philosophical Topics*, 14(2), p.5–31, 1986a.

RAILTON, P. "Moral Realism", in *Philosophical Review*, 95(2), p.163–207, 1986b.

RAWLS, J. "Outline of a Decision Procedure for Ethics", in *Philosophical Review*, 60(2), p.177–97, 1951.

RAWLS, J. **A Theory of Justice.** Cambridge, MA: Harvard University Press, 1971.

ROSATI, C. S. "Persons, Perspectives, and Full Information Accounts of the Good", in *Ethics*, 105(2), p.296–325, 1995.

ROSE, D., Nichols, S. "The Lesson of Bypassing", in *Review of Philosophy and Psychology*, 4(4), p.599–619, 2013.

ROSS, l., Nisbett, R. E. **The Person and the Situation**: Perspectives of Social Psychology. Philadelphia, Temple University Press, 1991.

RUSSELL, D. C. **Practical Intelligence and the Virtues**. Oxford: Oxford University Press, 2009.

RUSSELL, D. C. "From Personality to Character to Virtue", in Mark Alfano (ed.), *Current Controversies in Virtue Theory*, New York: Routledge, p. 91–106, 2015.

SAMUELS, R., STICH, S. “Rationality”, in *Encyclopedia of Cognitive Science*. Chichester: Wiley, 2002.

SAMUELS, S. M., CASEBEER, W.D. “A social psychological view of morality: Why knowledge of situational influences on behaviour can improve character development practices”, in *Journal of Moral Education*, 34(1), p.73–87, 2005.

SARKISSIAN, H. “Minor Tweaks, Major Payoffs: The Problems and Promise of Situationism in Moral Philosophy”, *Philosophers’ Imprint*, 10(9), <http://hdl.handle.net/2027/spo.3521354.0010.009>

SARKISSIAN, H., WRIGHT, J.C. (eds.). **Advances in Experimental Moral Psychology.** London: Bloomsbury Press, 2014.

SAYRE-MCCORD, G. “Introduction: The Many Moral Realisms”, in Sayre-McCord (1988b), p.1–24, 1988a.

SAYRE-MCCORD, G. (ed.) **Essays in Moral Realism.** Ithaca and London, Cornell University Press, 1988b.

SAYRE-MCCORD, G. “Moral Realism”, *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.)<https://plato.stanford.edu/archives/spr2015/entries/moral-realism/>

SCHNALL, s. et. al. “Disgust as Embodied Moral Judgment”, in *Personality and Social Psychology Bulletin*, 34(8), p.1069–1109, 2008a.

SCHNALL, S., BENTON, J., HARVEY, S. “With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments”, in *Psychological Science*, 19(12), p.1219–1222, 2008b.

SCHWITZGEBEL, E.; CUSHMAN, F. “Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers”, in *Mind and Language*, 27(2), p.135–153, 2011.

SCHWITZGEBEL, E.; CUSHMAN, F. “Philosophers’ Biased Judgments Persist Despite Training, Expertise, and Reflection”, in *Cognition*, 141, p.127–137, 2015.

SHAFER-LANDAU, R. *Moral Realism: A Defence*, Oxford: Clarendon Press, 2003.

SHERMAN, R. A., NAVE, C. S., FUNDER, D.C. “Situational Similarity and Personality Predict Behavioral Consistency”, in *Journal of Personality and Social Psychology*, 99(2), p. 330–343, 2010.

SHWEDER, R. A., BOURNE, E. J. “Does the Concept of the Person Vary Cross-Culturally?” in Anthony J. Marsella and Geoffrey M. White (eds.), *Cultural Conceptions of Mental Health and Therapy*. Boston, MA: Reidel Publishing, 1982.

SINGER, p. “Sidgwick and Reflective Equilibrium”, in *Monist*, 58(3), p.490–517, 1974.

SINNOTT-ARMSTRONG, W. P. “Moral Intuitionism Meets Empirical Psychology”, in Terry Horgan and Mark Timmons (eds.), *Metaethics After Moore*, New York: Oxford University Press, 2005.

SINNOTT-ARMSTRONG, W. P. (ed.) **Moral Psychology**, Vol. 1, The Evolution of Morality: Adaptations and Innateness. Cambridge, MA: MIT Press, 2008a.

SINNOTT-ARMSTRONG, W. P. (ed.) **Moral Psychology**, Vol. 2, The Cognitive Science of Morality: Intuition and Diversity. Cambridge, MA: MIT Press, 2008b.

SINNOTT-ARMSTRONG, W. P. (ed.) **Moral Psychology**, Vol.3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development. Cambridge, MA: MIT Press, 2008c.

SINNOTT-ARMSTRONG, W. P. (ed.) **Moral Psychology**, Vol.4, Free Will and Moral Responsibility. Cambridge, MA: MIT Press, 2014.

SLOTE, M. A. “Egoism and Emotion”, in *Philosophia*, 41(2), p. 313–335, 2013.

SMILANSKY, S. “Compatibilism: the Argument from Shallowness”, in *Philosophical Studies*, 115(3), p.257–282, 2003.

SMITH, A. **The Theory of Moral Sentiments.** London, Henry G. Bohn, 1853. [originalmente publicada em 1759]

SMITH, M. **The Moral Problem.** Cambridge, Basil Blackwell, 1994.

SNARE, F.E. “The Diversity of Morals”, in *Mind*, 89(355), p.353–369, 1980.

SNOW, N. E. **Virtue as Social Intelligence**: An Empirically Grounded Theory. London and New York, Routledge, 2010.

SOBER, E., WILSON, D.S. **Unto Others**: The Evolution and Psychology of Unselfish Behavior. Cambridge, MA: Harvard University Press, 1998.

SOLOMON, R. C. “Victims of Circumstances? A Defense of Virtue Ethics in Business”, in *Business Ethics Quarterly*, 13(1), p. 43–62, 2003.

SOLOMON, R. C. What’s Character Got to Do with It?’”, in *Philosophy and Phenomenological Research*, 71(3), p. 648–655, 2005.

SOSA, E. “Intuitions: Their Nature and Epistemic Efficacy”, in *Grazer Philosophische Studien*, 74(1), p. 51–67, 2007.

SOSA, E. “Situations Against Virtues: The Situationist Attack on Virtue Theory”, in Chrysostomos Mantzavinos (ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, New York: Cambridge University Press. P. 274–290, 2009.

SREENIVASAN, G. “Errors about errors: Virtue theory and trait attribution”, in *Mind*, 111(441), p. 47–68, 2002.

SRIPADA, C. S. “What Makes a Manipulated Agent Unfree?”, in *Philosophy and Phenomenological Research*, 85(3), p. 563–93, 2012.

STICH, S. **The Fragmentation of Reason**: Preface to a Pragmatic Theory of Cognitive Evaluation. Cambridge, MA: The MIT Press, 1990.

STICH, S, DORIS, J.M., ROEDDER, E.D. “Altruism”, in Doris et al. 2010, p.147–205, 2010.

STICH, S.; TOBIA, K. P. “Experimental Philosophy and the Philosophical Tradition”, in Sytsma and Buckwalter (2016), p.3–21, 2016.

STICH, S.; TOBIA, K. P. “Intuition and Its Critics”, in Stuart, Fehige, Brown (2018, cap. 21), 2018.

STICHTER, M. “Ethical Expertise: The Skill Model of Virtue”, in *Ethical Theory and Moral Practice*, 10(2), p.183–194, 2007.

STICHTER, M. “Virtues, Skills, and Right Action”, in *Ethical Theory and Moral Practice*, 14(1), p.73–86, 2011.

STRAWSON, P.F. “Freedom and Resentment”, in Gary Watson (ed.), *Free Will*, New York: Oxford University Press, 1982. [originalmente publicada em 1962].

STRAWSON, G. **Freedom and Belief.** Oxford: Oxford University Press, 1986.

STROHMINGER, N., LEWIS, R.L., MEYER, D. E. “Divergent Effects of Different Positive Emotions on Moral Judgment”, in *Cognition*, 119(2), p. 295–300, 2011.

STUART, M. T., FEHIGE, Y., BROWN, J.R. (eds.) **The Routledge Companion to Thought Experiments.** New York: Routledge, 2018.

STURGEON, N. L. “Moral Explanations”, in Sayre-McCord (1988b), p.229–255, 1988.

SUMNER, W. G. **Folkways**. Boston: Ginn and Company, 1934.

SUNSTEIN, C. R. “Moral Heuristics”, *Behavioral and Brain Sciences*, 28(4), p.531–42, 2005.

SWANTON, C. **Virtue Ethics**: A Pluralistic View. Oxford: Oxford University Press, 2003.

SYTSMA, J., BUCKWALTER, W. (eds.) **A Companion to Experimental Philosophy.** Oxford, Blackwell, 2016.

TETLOCK, P.E. “Review of *Culture of Honor: The Psychology of Violence in the South* by Robert Nisbett and Dov Cohen”, in *Political Psychology*, 20(1), p.211–13, 1999.

TIBERIUS, V. **Moral Psychology**: A Contemporary Introduction. New York: Routledge, 2015.

TOBIA, K. P., CHAPMAN, G.B., STICH, S. “Cleanliness is Next to Morality, Even for Philosophers”, in *Journal of Consciousness Studies*, 20(11 and 12): p.195–204, 2013.

TVERSKY, A.; KAHNEMAN, D. “Availability: A heuristic for judging frequency and probability”, in *Cognitive Psychology*, 5(2), p. 207–232, 1973.

TVERSKY, A.; KAHNEMAN, D. “The Framing of Decisions and the Psychology of Choice”, in *Science*, 211(4481), p.453–463, 1981.

UPTON, C. L. **Situational Traits of Character**: Dispositional Foundations and Implications for Moral Psychology and Friendship. Lanham, Maryland, Lexington Books, 2009.

VARGAS, M. “The Revisionist’s Guide to Responsibility”, in *Philosophical Studies*, 125(3), p.399–429, 2005a.

VARGAS, M. “Responsibility and the Aims of Theory: Strawson and Revisionism”, in *Pacific Philosophical Quarterly*, 85(2), p.218–241, 2005b.

VALDESOLO, P., DESTENO, D. “Manipulations of Emotional Context Shape Moral Judgment”, in *Psychological Science*, 17(6), p. 476–477, 2006.

VELLEMAN, J. d. “What Happens When Someone Acts?”, in *Mind*, 101(403), p.461–81, 1982.

VOYER, B. G., TARANTOLA, T. (eds.) **Moral Psychology**: A Multidisciplinary Guide. Springer [a ser publicado].

VRANAS, P.B.M. “The Indeterminacy Paradox: Character Evaluations and Human Psychology”, in *Noûs*, 39(1), p.1–42, 2005.

WATSON, G. “Two Faces of Responsibility”, in *Philosophical Topics* 24(2), p.227–48, 1996.

WEBBER, J. “Character, Consistency, and Classification”, in *Mind*, 115(459), p.651–658, 2006a.

WEBBER, J. “Virtue, Character and Situation”, in *Journal of Moral Philosophy*, 3(2), p.193–213, 2006b.

WEBBER, J. “Character, Common-Sense, and Expertise”, in *Ethical Theory and Moral Practice*, 10(1), p.89–104, 2007a.

WEBBER, J. “Character, Global and Local”, in *Utilitas*, 19(04), p.430–434, 2007b.

WESTERMARCK, E. *Origin and Development of the Moral Ideas*, 2 volumes, New York: MacMillian, 1906.

WIEGMANN, A., OKAN, Y., NAGEL, J. “Order Effects in Moral Judgment”, in *Philosophical Psychology*, 25(6), p.813–836, 2012.

WILLIAMS, B. “A Critique of Utilitarianism”, in *Utilitarianism: For and Against*, by J.J.C. Smart and Bernard Williams, Cambridge: Cambridge University Press, 1973.

WILLIAMS, B. **Ethics and the Limits of Philosophy**. Cambridge, MA: Harvard University Press, 1985.

WOOLFOLK, R. L., DORIS, j. m., Darley, J.M. “Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility”, in *Cognition*, 100(2), p.283–301, 2006.

ZHONG, C-B, STREJCEK, B., SIVANATHAN, N. “A Clean Self Can Render Harsh Moral Judgment”, in *Journal of Experimental Social Psychology*, 46(5), p.859–862, 2010.

ZIMBARDO, P. G. **The Lucifer Effect**: Understanding How Good People Turn Evil. Oxford: Blackwell Publishing Ltd, 2007.

(II) Raciocínio Moral* [39]

Autoria: Henry Richardson
Tradução: Rafael Martins
Revisão: Sagid Salles

---

* RICHARDSON, H. S. Moral Reasoning. In: ZALTA, E. (ed.) **The Stanford Encyclopedia of Philosophy**. Fall Edition, Stanford, CA: The Metaphysics Research Lab, Ago, 2018. Disponível em: https://plato.stanford.edu/archives/fall2018/entries/reasoning-moral/. Acesso em: 02 jul. 2022.

The following is the translation of the entry on Moral Reasoning by Henry Richardson, in the Stanford Encyclopedia of Philosophy. The translation follows the version of the entry in the SEP's archives at https://plato.stanford.edu/archives/fall2018/entries/reasoning-moral/
This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. 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.

[39] N.T.: A expressão benthamita refere-se a Jeremy Bentham.
Normative stringency: adstringência normativa refere-se a força ou peso de uma consideração normativa em competição com outras considerações e em admissões de exceções. Quanto mais adstringente, mais forte é a consideração, ou seja, ela normalmente vence outras considerações em deliberações assim como também menor a possibilidade de tal consideração admitir exceções.
A tradução da palavra *overriding* e suas variantes, *overrdingly, override, overridingness*, requer um esclarecimento, uma vez que o conceito de overridingness é muito utilizado em metaética. Ele se refere à propriedade que motivações, razões, ou juízos normativos, entre eles, os morais, possuem de superar, preponderar, ou prevalecer sobre motivações, razões, ou juízos de outras natureza, por exemplo, política, estética, prudencial, econômica, etc., em função de seu peso, grau, ou força normativa. As traduções para o verbo *to override* neste texto são: dominar, prevalecer, suplantar, superar, predominar. Para sua adjetivação ou nomeação, sugiro "suprema". (vide verbete **Brink Supremacy**, da SEF).
Agradecimento a Roseline Crippa pela exímia revisão.

Paradigmaticamente, o raciocínio moral é um raciocínio prático em primeira pessoa realizado por um agente individual ou coletivo sobre o que ele deve fazer moralmente, embora este tipo de raciocínio até possa ser realizado em nome de outra pessoa. O exame filosófico do raciocínio moral enfrenta dois enigmas distintos – um sobre como reconhecemos considerações morais e lidamos com os conflitos entre elas e sobre como eles nos movem a agir – e um sobre como adquirimos conhecimento sobre o que devemos fazer a partir de como raciocinamos sobre o que devemos fazer.

A Parte I deste artigo caracteriza o raciocínio moral de forma mais completa, situa-o em relação tanto a teorias de primeira ordem sobre a moralidade quanto a teorias metafísicas da moralidade e explica a significância do tópico. A Parte II então aborda uma série de questões filosóficas sobre o raciocínio moral, assim compreendido e situado.

1. A Importância Filosófica do Raciocínio Moral

1.1 Definindo "Raciocínio Moral"

Este artigo assume o raciocínio moral como uma espécie de raciocínio prático - isto é, como um tipo de raciocínio voltado a decidir sobre o que fazer e, quando bem-sucedido, gerar uma intenção (ver verbete sobre razão prática). Claro, também raciocinamos teoricamente sobre o que a moralidade exige de nós; mas a natureza do raciocínio puramente teórico sobre a ética é abordada de forma adequada nos vários artigos sobre ética. Também é verdade que, em alguns entendimentos, o raciocínio moral voltado para decidir o que fazer envolve formar juízos sobre o que se deve, moralmente, fazer. Nessas compreensões, perguntar o que se deve (moralmente) fazer pode ser uma questão prática, uma certa maneira de perguntar sobre o que fazer. (Veja a seção 1.5 sobre a questão de saber se esta é uma questão prática distinta.) Para fazer justiça a toda a gama de visões filosóficas sobre o raciocínio moral, precisaremos ter uma compreensão ampla do que conta como uma questão moral. Por exemplo, visto que uma posição proeminente sobre o raciocínio moral é que as considerações relevantes não são codificáveis, cometeríamos uma petição de princípio central se aqui definíssemos "moralidade" como envolvendo princípios ou regras codificáveis. Para os presentes propósitos,

podemos entender discussões sobre o que é certo ou errado, virtuoso ou vicioso, como levantando questões morais.

Mesmo quando questões morais surgem explicitamente na vida diária, como quando nos deparamos com questões sobre educação de filhos, questões agrícolas e de negócios, às vezes agimos impulsivamente ou instintivamente, em vez de parar para raciocinar, não apenas sobre o que fazer, mas sobre o que nós devemos fazer. Jean-Paul Sartre descreve o caso de um de seus alunos que o procurou na Paris ocupada durante a Segunda Guerra Mundial, pedindo conselhos sobre se deveria juntar-se às forças da França Livre, então se concentrando na Inglaterra ou se deveria ficar com sua mãe, que do contrário ficaria sozinha (SARTRE, 1975). No amplo sentido que acabamos de descrever, esta é provavelmente uma questão moral; e o jovem hesitou por tempo suficiente para pedir o conselho de Sartre. Isso significa que esse jovem está raciocinando sobre sua questão prática? Não necessariamente. Na verdade, Sartre usou o caso para expor seu ceticismo sobre a possibilidade de abordar essa questão prática por meio do raciocínio. Mas o que é raciocínio?

O raciocínio, do tipo discutido aqui, é o pensamento ativo ou explícito, em que aquele que raciocina tenta alcançar uma resposta bem sustentada a uma pergunta bem definida (HIERONYMI, 2013), responsavelmente guiado por suas avaliações de suas razões (KOLODNY, 2005) e por quaisquer obrigações racionais aplicáveis (BROOME, 2009, 2013). Para o aluno de Sartre, pelo menos essa questão surgiu. De fato, a questão era relativamente definida, implicando que o aluno já havia se engajado em alguma reflexão sobre as várias alternativas disponíveis para ele - um processo que tem sido bem descrito como uma fase importante do raciocínio prático, que precede apropriadamente o esforço de se tomar uma decisão (HARMAN, 1986, p. 2).

É claro que caracterizar o raciocínio como um pensamento conduzido de forma responsável não é suficiente para analisar o conceito. Em primeiro lugar, tal caracterização falha em abordar a intrincada questão sobre a relação do raciocínio com a inferência (HARMAN, 1986, BROOME, 2009). Além disso, não se resolve a questão sobre se formular uma intenção sobre o que fazer é suficiente para concluir o raciocínio prático ou se tais intenções não podem ser adequadamente elaboradas até que se comece a agir. Talvez não se possa raciocinar adequadamente sobre como consertar uma parede de tijolos ou como fazer uma omelete com os ingredientes disponíveis sem realmente começar a consertar ou cozinhar (*vide* FERNANDEZ, 2016). Ainda assim, esta caracterização servirá para os propósitos presentes. Basta

deixar claro que a noção de raciocínio envolve normas de pensamento. Essas normas de aptidão ou correção do pensamento prático certamente não exigem que pensemos ao longo de um único caminho prescrito, mas permitem apenas certos caminhos e não outros (BROOME, 2013, p. 219). Mesmo assim, sem dúvida, muitas vezes deixamos de cumpri-los.

1.2 Desafios Empíricos ao Raciocínio Moral

Nosso pensamento, incluindo nosso pensamento moral, muitas vezes não é explícito. Poderíamos dizer que também raciocinamos tacitamente, pensando da mesma forma que no raciocínio explícito, mas sem qualquer tentativa explícita de chegar a respostas bem fundamentadas. Em algumas situações, inclusive as morais, podemos ser imprudentes se tentarmos responder nossas questões práticas por meio de raciocínio explícito. Em outras situações, pode até ser um erro raciocinar tacitamente - porque, digamos, enfrentamos uma emergência iminente. “Às vezes não devemos deliberar sobre o que fazer, e apenas agir” (ARPALY; SCHROEDER, 2014, p. 50). No entanto, mesmo que não sejamos chamados a pensar sobre nossas opções em todas as situações, e mesmo se às vezes seria positivamente melhor se não o fizéssemos, ainda assim, se somos chamados a fazê-lo, então devemos conduzir nosso pensamento com responsabilidade: devemos raciocinar.

Trabalhos recentes em ética empírica indicam que, mesmo quando somos convocados a raciocinar moralmente, frequentemente o fazemos muito mal. Quando somos solicitados a oferecer razões para nossas intuições morais, muitas vezes ficamos “perplexos”, não encontrando nada a dizer em defesa de tais intuições (HAIDT, 2001). Nosso pensamento sobre cenários morais hipotéticos tem demonstrado ser altamente sensível a variações arbitrárias, como ordem de apresentação. Descobriu-se que mesmo filósofos profissionais são propensos a tais lapsos de raciocínio claro (por exemplo, SCHWITZGEBEL; CUSHMAN, 2012). Parte de nossa perplexidade e confusão é atribuída ao fato de termos, ao mesmo tempo, uma maneira rápida e mais emocional de processar estímulos morais e uma forma lenta e mais cognitiva (por exemplo, GREENE, 2014). Uma explicação alternativa da perplexidade moral olha para as normas sociais de raciocínio moral (SNEDOON, 2007). E uma reação mais otimista à nossa confusão vê nossos padrões estabelecidos de "raciocínio sobre consistência moral" como sendo adequados para lidar com a percepção conflitante gerada por nossos sistemas rápidos e lentos (CAMPBELL;

KUMAR, 2012) ou como constituindo "um sistema de aprendizagem flexível que gera e atualiza uma estrutura avaliativa multidimensional para orientar a decisão e a ação" (RAILTON, 2014, p. 813).

Eventualmente, esse trabalho empírico sobre nosso raciocínio moral pode resultar em revisões em nossas normas de raciocínio moral. Isso ainda não aconteceu. Este artigo ocupa-se principalmente com questões filosóficas colocadas por nossas normas atuais de raciocínio moral. Por exemplo, dadas essas normas e supondo que sejam mais ou menos seguidas, como as considerações morais fatoram no raciocínio moral, como são classificadas pela razão quando se chocam, e como levam à ação? E o que essas normas indicam sobre o que devemos fazer?

1.3 Situando o Raciocínio Moral

O tópico do raciocínio moral está entre dois outros tópicos comumente abordados na filosofia moral. Por um lado, há a questão de primeira ordem sobre quais verdades morais existem, se é que há verdades morais. Por exemplo, existem princípios morais gerais verdadeiros e, em caso afirmativo, quais são eles? Nesse nível, o utilitarismo compete com o kantismo, por exemplo, e ambos competem com antiteóricos de várias tendências, que reconhecem apenas verdades particulares sobre a moralidade (CLARKE; SIMPSON, 1989). Por outro lado, um tipo bastante diferente de questão surge da tentativa de fornecer uma base metafísica para verdades morais ou para a afirmação de que não existem. Supondo que existam algumas verdades morais, o que as **torna** verdadeiras? Qual explicação pode ser dada sobre as condições de verdade das declarações morais? Aqui surgem questões familiares ao **ceticismo moral** e **relativismo moral**; aqui, a ideia de "uma razão" é empenhada por muitos na esperança de defender uma metafísica moral não-cética (*vide* SMITH, 2013). O tópico do raciocínio moral está entre esses dois outros tópicos familiares no seguinte e simples sentido: os raciocinadores morais operam com o que consideram ser moralmente verdadeiro, mas, em vez de perguntar o que "*torna*" suas crenças morais verdadeiras, eles procedem **responsavelmente para tentar descobrir** o que fazer à luz dessas considerações. O estudo filosófico do raciocínio moral se ocupa com a natureza dessas tentativas.

Esses três tópicos estão claramente relacionados. É concebível que as relações entre eles sejam tão estreitas a ponto de excluir qualquer interesse independente no tópico do raciocínio moral. Por exemplo, se tudo o que poderia

ser dito de maneira útil sobre o raciocínio moral fosse que ele se resume a atentar para os fatos morais, então todo o interesse recairia sobre a questão do que são esses fatos - com algum foco residual na ideia de percepção moral (MCNAUGHTON, 1988). Alternativamente, pode-se pensar que o raciocínio moral é simplesmente uma questão de se aplicar a teoria moral correta por meio dos modos comuns de raciocínio dedutivo e empírico. Novamente, se isso fosse verdade, o objetivo suficiente para qualquer pessoa seria encontrar essa teoria e instruir-se dos fatos não-morais corretos. Entretanto, nenhum desses extremos redutivos parece plausível. Consideremos primeiro a potencial redução do raciocínio moral a uma interpretação correta dos fatos.

Defensores contemporâneos da importância de percebermos corretamente os fatos moralmente relevantes tendem a se concentrar em fatos que podemos perceber usando nossas faculdades sensíveis comuns e nossas capacidades comuns de reconhecimento, como "*que essa pessoa tem uma infecção*" ou *"que precisa da minha ajuda médica"*. Nesta base, é possível lançar argumentos poderosos contra a afirmação de que os princípios morais sustentam cada verdade moral (DANCY, 1993) e para a tese de que às vezes podemos perfeitamente decidir o que fazer agindo com base nas razões que percebemos instintivamente – ou em como fomos treinados - sem nos empenhar em qualquer raciocínio moral. No entanto, esta **não** é uma base sólida para argumentar que o raciocínio moral, no sentido que vai além de simplesmente se atentar aos fatos morais, é sempre desnecessário. Pelo contrário, muitas vezes nos encontramos enfrentando novas perplexidades e conflitos morais para os quais nossa percepção moral é um guia inadequado. Por exemplo, ao abordar as questões morais sobre se a sociedade deve permitir contratos de maternidade substituta, as novidades científicas e tecnológicas envolvidas tornam nossas percepções morais em guias duvidosos e incertos. Quando um médico pesquisador observando a doença de um paciente também nota que desviar recursos para cuidar, clinicamente, desse indivíduo inibiria o progresso de sua pesquisa, prejudicando assim as chances de cura a longo prazo de futuros portadores dessa doença, o médico pesquisador fica cara a cara com considerações morais conflitantes. Nesta dicotomia, ponderando ambas as considerações, é muito menos plausível ou satisfatório dizer que, empregando nossas capacidades sensoriais e de reconhecimento, simplesmente veremos o que deve ser feito. Postular uma faculdade especial de intuição moral capaz de fazer tais julgamentos gerais em face de considerações conflitantes equivale a introduzir um "*deus ex machina*". Este movimento encerra a investigação de uma forma que

atende melhor aos propósitos de uma ficção do que aos propósitos de compreensão. Em vez disso, é mais plausível supor que o raciocínio moral entra exatamente neste ponto (CAMPBELL; KUMAR, 2012).

Para os presentes propósitos, vale a pena notar, David Hume e os teóricos do senso moral não contam como obstruindo nosso entendimento do raciocínio moral dessa maneira. É verdade que Hume se apresenta, especialmente no *Tratado da Natureza Humana*, como um descrente de qualquer raciocínio especificamente prático ou moral. Ao fazer isso, no entanto, ele emprega uma definição extremamente restrita de "raciocínio" (HUME, 2000, Livro I, Parte III, seção II). Para os presentes propósitos, em contraste, estamos usando uma abordagem mais ampla de "raciocínio", não limitada pela ambição de analisar as contribuições relativas da (faculdade da) razão e das paixões. E sobre o raciocínio moral neste sentido mais amplo, como pensamento responsável sobre o que se deve fazer, Hume tem muitas coisas interessantes a dizer, começando com o pensamento de que o raciocínio moral deve envolver uma dupla correção de perspectiva (*vide* seção 2.4) de modo a fazer jus as reivindicações de outras pessoas e de reconhecer um futuro mais distante, uma dupla checagem que se realiza com o auxílio das chamadas "paixões calmas".

Haverá razões para permanecermos céticos se passarmos da possibilidade de que compreender os fatos corretamente substituiria o raciocínio moral para a possibilidade de que a aplicação da teoria moral correta resumiria - ou exauriria - o raciocínio moral. Uma das razões é que as teorias morais não surgem no vácuo; em vez disso, elas se desenvolvem dentro de uma ampla estrutura de convicções morais. Na medida em que é forte a primeira vertente potencialmente redutiva, enfatizando a importância de perceber os fatos morais – e ela tem alguma força - ela também tende a mostrar que as teorias morais precisam ganhar sustentação através da sistematização ou explicação de uma ampla gama de fatos morais (SIDGWICK, 1981). Como na maioria das outras áreas em que a explicação teórica é necessária, o grau de sucesso explicativo permanecerá parcial e aberto a melhorias por meio de revisões na teoria (*vide* seção 2.6). No entanto, ao contrário das ciências naturais, a teoria moral é um empreendimento "socrático" no sentido de que é um assunto pertencente a ações "moldadas pelo autoexame", como John Rawls certa vez colocou (RAWLS, 1971, p. 48-.). Se essa observação estiver correta, sugere que as questões morais que nos propomos a responder surgem de nossas reflexões sobre aquilo que importa. Da mesma forma - e este é o ponto aqui - uma teoria moral está sujeita a ser derrubada porque gera implicações concretas que não nos agradam sob a devida reflexão. Sendo assim, e admitindo a grande complexidade

do terreno moral, parece altamente improvável que algum dia chegaremos a uma teoria moral com base na qual possamos progredir de forma dedutiva com serenidade e confiança na busca por respostas ao que devemos fazer em todos os casos concretos. Esta conclusão é reforçada por uma segunda consideração, a saber, que na medida em que uma teoria moral é fiel à complexidade dos fenômenos morais, ela conterá em si mesma muitas possibilidades de conflitos entre seus próprios elementos. Mesmo que ela inclua algumas regras de prioridade, é improvável que estas regras sejam capazes de cobrir todas as contingências. Consequentemente, é inevitável que algum raciocínio moral que vá além da aplicação dedutiva da teoria correta seja necessário.

Em suma, um entendimento sólido do raciocínio moral não tomará a forma de reduzi-lo a um dos outros dois níveis de filosofia moral identificados acima. Nem a exigência de atender aos fatos morais, nem a diretiva de aplicar a teoria moral correta exaurem ou descrevem suficientemente o raciocínio moral.

1.4. Obtendo Conhecimento Moral a partir do Estudo do Raciocínio Moral

Além de apresentar problemas filosóficos por si só, o raciocínio moral é interessante por causa de suas implicações para os fatos morais e teorias morais. Consequentemente, atentar-se ao raciocínio moral frequentemente será útil para aqueles cujo real interesse é determinar a resposta correta para algum problema moral concreto ou argumentar a favor ou contra alguma teoria moral. Os métodos característicos com que tentamos resolver um determinado tipo de dilema moral podem ser tão reveladores sobre nossas abordagens ponderadas dessas questões quanto qualquer juízo final a que possamos caracteristicamente chegar por meio desses métodos. Além disso, podemos ter convicções refletidas e firmes sobre a melhor maneira de lidar deliberadamente com uma determinada classe de problemas, mesmo quando continuamos em dúvida sobre o que deve ser feito. Em tais casos, atentar para os modos de raciocínio moral que caracteristicamente aceitamos pode expandir de forma útil o conjunto de informações morais a partir do qual nosso raciocínio avança, sugerindo maneiras de estruturar as considerações concorrentes.

Fatos sobre a natureza da inferência moral e do raciocínio moral podem ter implicações diretas significativas para a teoria moral. Por exemplo, pode ser considerado uma condição de adequação para qualquer teoria moral que ela desempenhe um papel realmente prático em nossos esforços de autocompreensão

e deliberação. Ela deve ser um guia de deliberação (RICHARDSON, 2018, §1.2). Se essa condição for aceita, então qualquer teoria moral que exigir que os agentes se engajem em raciocínios abstrusos ou especiosos pode ser inadequada por esse motivo, como o seria qualquer teoria que presumisse que indivíduos comuns geralmente são incapazes de raciocinar das maneiras que a teoria exige. J.S. Mill (1979) admite que geralmente somos incapazes de fazer os cálculos exigidos pelo utilitarismo, como ele o entendia, e argumentou que devemos nos consolar com o fato de que, ao longo da história, a experiência gerou princípios secundários que nos direcionam bem o suficiente. De forma bem mais dramática, R. M. Hare defendeu que o utilitarismo captava bem o raciocínio de "arcanjos" racionais e idealmente informados (1981). Levar a sério este critério de orientação de deliberação para teorias morais favoreceria, em vez disso, teorias que informam mais diretamente os esforços de raciocínio moral feitos por nós "proletários", para usar o termo contrastante de Hare.

Consequentemente, as relações estreitas entre o raciocínio moral, os fatos morais e as teorias morais não eliminam o raciocínio moral como um tópico de interesse. Ao contrário, porque o raciocínio moral tem implicações importantes sobre fatos morais e teorias morais, essas relações íntimas geram interesse adicional ao tópico do raciocínio moral.

1.5. Quão Distinto é o Raciocínio Moral do Raciocínio Prático em Geral?

A questão básica final é se o raciocínio moral é verdadeiramente distinto do raciocínio prático compreendido de maneira mais geral. (A questão sobre saber se o raciocínio moral, mesmo que prático, é estruturalmente distinto do raciocínio teórico que simplesmente procede de um reconhecimento adequado dos fatos morais já foi implicitamente abordada e respondida, para os fins da presente discussão, de forma afirmativa). Ao abordar essa questão final, é difícil ignorar a maneira como diferentes teorias morais projetam modelos bastante diferentes de raciocínio moral - novamente um elo que pode ser buscado pelo filósofo moral que busca alavancagem em qualquer direção. Por exemplo, as teses de Aristóteles podem ser as seguintes: uma explicação bastante geral pode ser dada do raciocínio prático, que inclui a seleção de meios para os fins e a determinação dos constituintes de uma atividade desejada. A diferença entre o raciocínio de uma pessoa viciosa e o de uma pessoa virtuosa não reside em sua estrutura, mas apenas em seu

conteúdo, pois a pessoa virtuosa busca os verdadeiros bens, enquanto a pessoa viciosa simplesmente é desencaminhada pelos bens aparentes. Para esclarecer, a pessoa virtuosa pode conseguir uma maior integração de seus fins por meio do raciocínio prático (por causa da forma como as várias virtudes são coerentes entre si), mas esta é uma diferença no resultado do raciocínio prático e não em sua estrutura. Em um extremo oposto, o imperativo categórico de Kant tem sido entendido como gerando uma abordagem para o raciocínio prático (por meio de um "tipo de juízo prático") que é distinto de outros raciocínios práticos, tanto na variedade de considerações que aborda quanto em sua estrutura (NELL, 1975). Enquanto o raciocínio prático prudencial, na visão de Kant, visa maximizar a felicidade do agente, o raciocínio moral aborda a potencial universalização das máximas - grosso modo, as intenções - sobre as quais se age. As visões intermediárias entre Aristóteles e Kant a esse respeito incluem a teoria utilitarista de Hare e a teoria da lei natural de Tomás de Aquino. Do ponto de vista de Hare, assim como um agente prudencial ideal aplica a racionalidade de maximização às suas próprias preferências, o raciocínio de um agente moral ideal aplica a racionalidade de maximização ao conjunto de preferências de todos aqueles reconhecidos pela capacidade arcangélica de simpatia deste agente moral (HARE, 1981). As visões tomistas da lei natural compartilham a visão aristotélica sobre a unidade geral do raciocínio prático em busca do bom, bem ou mal concebido, e não obstante acrescentam que essa razão prática, não só exige que busquemos os bens humanos fundamentais, mas também, e notoriamente, exige que não vituperemos esses bens. Desta forma, as teorias da lei natural incorporam, dentro de uma teoria unificada do raciocínio prático, alguma estruturação distintamente moral - como a distinção entre a execução de uma ação e sua mera permissão, assim como a distinção entre intencionar algo como um meio e aceitar algo como efeito colateral, oriundas da chamada doutrina do duplo efeito (ver verbete sobre a tradição da lei natural na ética). À luz desta diversidade de pontos de vista sobre a relação entre o raciocínio moral e o raciocínio prático ou prudencial, uma explicação geral do raciocínio moral que se preocupa em evitar presumir a correção de uma teoria moral específica fará bem em permanecer agnóstica com respeito à questão sobre como o raciocínio moral se relaciona com o raciocínio prático não-moral.

2. Questões Filosóficas Gerais Sobre Raciocínio Moral

Certamente, a maioria dos grandes filósofos que abordaram a natureza do raciocínio moral estavam longe de ser agnósticos sobre o conteúdo da teoria moral correta e desenvolveram suas reflexões sobre o raciocínio moral em apoio ou em consequência de sua teoria moral. No entanto, discussões contemporâneas que se apresentam em algum grau agnósticas sobre o conteúdo das teorias morais têm levantado aspectos importantes e controversos do raciocínio moral. Podemos agrupá-los em torno das sete perguntas a seguir:

1. Como as considerações relevantes são abordadas no raciocínio moral?
2. Seria essencial para o raciocínio moral que as considerações que ele aborda sejam cristalizadas em, ou agrupadas sob, princípios?
3. Como classificamos quais considerações morais são mais relevantes?
4. De que maneiras os elementos motivacionais moldam o raciocínio moral?
5. Como melhor modelar os tipos de conflitos entre as considerações que surgem no raciocínio moral?
6. O raciocínio moral inclui aprendizado empírico e mudança de convicção?
7. Como podemos raciocinar, moralmente, no âmbito dialético?

O restante deste artigo aborda cada uma dessas sete questões.

2.1. Abordagem Moral

Uma vantagem de definir “raciocínio” de forma ampla, como feito aqui, é que ajuda a reconhecer que os processos pelos quais chegamos a estar concretamente conscientes sobre questões morais são parte integrante do raciocínio moral, entendido de forma mais restrita. Reconhecer questões morais quando elas surgem requer um conjunto de capacidades obtidas através de muito treino e uma sensibilidade emocional amplamente diversificada. Filósofos da escola do senso moral dos séculos

XVII e XVIII enfatizavam propensões emocionais inatas, como a simpatia por outros humanos. Os teóricos da virtude de influência clássica, em contraste, dão mais importância ao treinamento da percepção e ao amadurecimento emocional que deve acompanhá-lo. Entre os filósofos contemporâneos que trabalham em ética empírica há uma divisão semelhante, com alguns argumentando que processamos situações usando uma gramática moral inata (MIKHAIL, 2011) e alguns enfatizando o papel das emoções nesse processamento (HAIDT, 2001; PRINZ, 2007; GREENE, 2014). Pois ao raciocinar moralmente, uma tarefa crucial para nossas capacidades de reconhecimento moral é destacar certas características de uma situação como sendo moralmente relevantes. Por exemplo, o aluno de Sartre concentrou-se nas obrigações concorrentes relacionadas a sua mãe e a França Livre, atribuindo a cada consideração uma importância relativa à sua própria condição, que ele não atribui a comer queijo francês ou vestir uniforme. Dizer que certas características são destacadas como moralmente relevantes não significa implicar que as características assim isoladas respondam aos termos de algum princípio geral ou outro: chegaremos à questão do particularismo, abaixo. Em vez disso, é simplesmente sugerir que a capacidade de reconhecimento parece ter um foco seletivo.

Aquilo que conta como uma questão ou problema moral, no sentido de exigir o reconhecimento dos agentes morais, novamente variará de acordo com a teoria moral. Nem todas as teorias morais considerariam a lealdade filial e o patriotismo como deveres morais. É apenas a um grande custo, no entanto, que qualquer teoria moral poderia alegar prescindir de uma camada de pensamento moral envolvendo reconhecimento situacional. Talvez, possamos imaginar algum tipo de cálculo utilitarista segundo o qual não há necessidade de identificar uma questão ou problema moral, pois **todas** as escolhas na vida apresentam ao agente a mesma tarefa de maximização de utilidade. Talvez Jeremy Bentham tenha defendido um utilitarismo desse tipo. Para os utilitarismos mais plausíveis mencionados acima, no entanto, como os de Mill e Hare, os agentes nem sempre precisam calcular tudo novamente desde o início, mas devem estar atentos à possibilidade de precisar recorrer a um modo mais direto e crítico de raciocínio moral, porque os "pontos de referência e direção" mais comuns podem levar qualquer um ao erro na situação em questão. Reconhecer se alguém está em uma dessas situações torna-se assim a principal tarefa de reconhecimento para o agente utilitarista. Se essa tarefa pode ser adequadamente circunscrita, é claro, tem sido uma das questões cruciais sobre se tais formas indiretas de utilitarismo, atraentes por outros motivos, podem evitar ser equiparadas a uma forma mais direta e benthamita (*vide* BRANDT, 1979)

Observe que, ao descrever a abordagem moral, não sugerimos que o que é percebido é um fato moral. Em vez disso, pode ser que o que é percebido seja alguma característica comum e descritiva de uma situação que seja, por qualquer motivo, moralmente relevante. No entanto, uma explicação da abordagem moral colidirá de maneira interessante com a metafísica dos fatos morais, se sustentar que os fatos morais podem ser percebidos. Nesse aspecto, um importante intermediário é o conjunto de julgamentos envolvendo os chamados conceitos avaliativos "densos" – por exemplo, que alguém é insensível, grosseiro, justo ou corajoso (veja o verbete sobre conceitos éticos densos). Estes não invocam os termos supostamente "mais lacunares" da avaliação moral geral, "bom" ou "correto". No entanto, eles também não são inocentes de conteúdo normativo. Evidentemente, reconhecemos a insensibilidade quando vemos casos claros dela. Também evidentemente – quaisquer que sejam as implicações metafísicas do último fato – nossa capacidade de descrever nossas situações nesses termos normativos densos é crucial para nossa capacidade de raciocinar moralmente.

É debatido até que ponto nossas habilidades de discernimento moral estão ligadas às nossas motivações morais. Para Aristóteles e muitos de seus antigos sucessores, os dois estão intimamente ligados, pois alguém não educado em motivações virtuosas não verá as coisas corretamente. Por exemplo, os covardes superestimarão os perigos, os imprudentes os subestimarão e os virtuosos os perceberão corretamente (Ética Eudemica 1229b 23-27). Também pelos estoicos, ter as motivações corretas era considerado algo intimamente ligado à percepção correta do mundo; mas enquanto Aristóteles via as emoções como aliadas para apoiar o bom discernimento moral, os estoicos as viam como hostis à percepção clara da verdade (*vide* NUSSBAUM, 2001).

2.2. Princípios Morais

O fato de alguém discernir características e qualidades de alguma situação que são relevantes para avaliá-la moralmente ainda não implica que alguém explicitamente ou mesmo implicitamente empregue quaisquer afirmações gerais ao descrevê-la. Talvez tudo o que se perceba sejam características e qualidades particularmente incorporadas, sem percebê-las de forma saliente como instanciações de qualquer tipo. O aluno de Sartre pode estar focado em sua mãe e nas dificuldades particulares de vários de seus colegas franceses sob ocupação nazista, em vez de

em quaisquer supostas exigências de dever filial ou patriotismo. Tendo se tornado consciente de alguma questão moral em termos relativamente particulares, ele pode prosseguir diretamente para resolver o conflito entre eles. Outra possibilidade, no entanto, e que frequentemente parecemos explorar, é formular a questão em termos gerais: "Um filho único deve cuidar de uma mãe isolada", por exemplo, ou "deve-se ajudar aqueles em extrema necessidade se pode fazê-lo sem sacrifício pessoal significativo". Tais afirmações gerais seriam exemplos de "princípios morais", em sentido amplo. (Não fazemos distinção aqui entre princípios e regras. Entre os que fazem estão DWORKIN, 1978 e GERT, 1998.)

Devemos ser cuidadosos, aqui, para distinguir a questão sobre se os princípios geralmente desempenham um papel implícito ou explícito no raciocínio moral, incluindo o raciocínio moral bem conduzido, da questão sobre se os princípios necessariamente figuram como parte da base da verdade moral. A última questão é melhor entendida como uma questão metafísica sobre a natureza e a base dos fatos morais. O que atualmente é conhecido como particularismo moral é a tese de que não há princípios morais defensáveis e que razões morais, ou fatos morais bem fundamentados, podem existir independentemente de qualquer base em um princípio geral. Uma teoria contrária sustenta que as razões morais são necessariamente gerais, seja porque as fontes de sua justificação são todas gerais ou porque uma afirmação candidata a princípio moral é mal formulada se contém particularidades. Porém, se os princípios desempenham um papel útil no raciocínio moral é certamente uma questão diferente de se os princípios desempenham um papel necessário na explicação das condições de verdade últimas das declarações morais. O particularismo moral, como acabamos de definir, nega esta última função. Alguns particularistas morais parecem também acreditar que o particularismo moral **implica** que os princípios morais não podem desempenhar uma função útil no raciocínio. Essa alegação é discutível, pois parece uma questão contingente se os fatos particulares relevantes se organizam de maneira passível de sumarização geral e se nosso aparato cognitivo pode lidar com eles sem empregar princípios gerais. Embora a controvérsia metafísica sobre o particularismo moral esteja em grande parte fora do nosso tópico, vamos revisitá-la na seção 2.5, em conexão com a ponderação de razões conflitantes.

No que diz respeito ao raciocínio moral, embora existam alguns autodenominados "antiteóricos" que negam que estruturas abstratas de generalidades vinculadas sejam importantes para o raciocínio moral (CLARKE, et al. 1989), é mais comum encontrar filósofos que reconhecem tanto uma função para os julgamentos

particulares quanto para princípios morais. Assim, neoaristotélicos como Nussbaum, que enfatizam a importância do discernimento particular "refinado e ricamente consciente", também consideram que esse discernimento é guiado por um conjunto de virtudes descritíveis em formas gerais, e que essas descrições gerais entrarão em jogo em pelo menos alguns tipos de casos (NUSSBAUM, 1990). Os "situacionistas morais" de uma geração anterior (*vide* FLETCHER, 1997) enfatizaram a importância de levar em conta uma ampla variedade de diferenças circunstanciais, mas tendo como pano de fundo alguns princípios gerais cuja aplicação é auxiliada por aquelas diferenças. As feministas morais influenciadas pelo trabalho pioneiro de Carol Gilligan sobre desenvolvimento moral enfatizam a centralidade moral do tipo de cuidado e discernimento que são salientes e bem desenvolvidos em pessoas imersas em relacionamentos particulares (HELD, 1995); mas essa ênfase é consistente com princípios gerais como, por exemplo, "devemos ser sensíveis aos desejos de nossos amigos" (*vide* verbete sobre **psicologia moral feminista**). Mais uma vez, se distinguirmos a questão de saber se os princípios são úteis no pensamento moral conduzido de forma responsável da questão sobre se as razões morais, em última análise, derivam de princípios gerais, e concentrarmos nossa atenção apenas na primeira, veremos que algumas das oposições a princípios morais gerais se desfazem.

Deve-se notar que estamos usando uma noção fraca de generalidade, aqui. Ela contrasta apenas com o tipo de particularidade estrita que acompanha indexicais e nomes próprios. Afirmações gerais – aquelas que não contêm tais referências particulares – não são necessariamente generalizações universais, fazendo uma afirmação sobre **todos** os casos do tipo mencionado. Assim, "normalmente deve-se ajudar aqueles em extrema necessidade" é um princípio geral, neste sentido fraco. Possivelmente, tais princípios logicamente fracos seriam ofuscantes no contexto de uma tentativa de reconstruir as condições últimas de verdade das declarações morais. Tais princípios logicamente fracos seriam claramente inúteis em qualquer tentativa de gerar um "silogismo prático" dedutivamente estrito. Em nosso raciocínio não dedutivo do dia-a-dia, no entanto, esses princípios logicamente fracos parecem ser bastante úteis. (Lembre-se de que estamos entendendo "raciocínio" de forma bastante ampla, como pensamento conduzido com responsabilidade: nada neste entendimento do raciocínio sugere qualquer lugar privilegiado para a inferência dedutiva: *vide* HARMAN, 1986. Para mais informações sobre princípios falíveis ou "padrão", consulte a seção 2.5.)

Nesta terminologia, estabelecer que os princípios gerais são essenciais para o raciocínio moral deixa em aberto a questão sobre se os princípios logicamente

estritos ou que não admitem exceções também são essenciais para o raciocínio moral. Certamente, muito do nosso raciocínio moral real parece ser impulsionado por tentativas de reformular ou reinterpretar princípios para que possam **ser considerados** como não admitindo exceções. Adeptos e herdeiros da tradição da lei natural na ética (por exemplo, DONAGAN, 1977) são defensores particularmente elegantes de princípios morais estritos, pois são capazes de valer-se não apenas de uma tradição casuística refinada, mas também de uma ampla gama de sutis - alguns diriam excessivamente sutis – distinções, como as mencionadas acima, entre execução e mera permissão, e entre intencionar como meio e aceitar como subproduto.

Uma função relacionada a uma versão forte de generalidade no raciocínio moral vem do pensamento kantiano de que o raciocínio moral deve se opor à tendência do agente de fazer exceções convenientes para si mesmo. Como observamos, Kant sustenta que devemos perguntar se as máximas de nossas ações podem servir como leis universais. De acordo com o modo como a maioria dos leitores contemporâneos entende essa demanda, ela exige que nos engajemos em uma espécie de generalização hipotética entre os agentes e perguntemos sobre as implicações de todos agirem dessa maneira nessas circunstâncias. Os fundamentos para desenvolver o pensamento de Kant nessa direção foram bem explorados (*vide* NELL, 1975, KORSGAARD, 1996, ENGSTROM, 2009). A importância e as dificuldades de tal teste hipotético de generalização em ética foram discutidas nos influentes trabalhos Gibbard (1965) e Goldman (1974).

2.3 Selecionando Quais Considerações São Mais Relevantes

Independentemente de considerações morais precisarem ou não do respaldo de princípios gerais, devemos esperar que os contextos agenciais nos apresentem múltiplas considerações morais. Além disso, é claro, essas situações também nos apresentarão muitas informações que não são moralmente relevantes. Em qualquer explicação realista, uma tarefa central do raciocínio moral é separar as considerações relevantes das irrelevantes, bem como determinar quais são especialmente relevantes e quais apenas ligeiramente. Que uma certa mulher é a **mãe** do aluno de Sartre parece ser um fato moralmente relevante; e quanto ao fato (supondo que seja um) que ela não tem outros filhos para cuidar dela? Abordando a tarefa de separar o que é moralmente relevante do que não é, alguns filósofos oferecem explicações gerais de características moralmente relevantes. Outros buscam explicar como

classificamos quais das características relevantes são **mais** relevantes, um processo de pensamento que às vezes é chamado de "casuística".

Antes de olharmos para as maneiras de classificar quais características são moralmente relevantes ou **mais** moralmente relevantes, pode ser útil observar um passo anterior dado por alguns casuístas, que é tentar estabelecer um esquema que capture **todas** as características de uma ação ou proposta de ação. Os casuístas católicos romanos da Idade Média fizeram isso baseando-se nas categorias de Aristóteles. Assim, eles perguntaram onde, quando, por que, como, por que meios, para quem e por quem a ação em questão deve ser feita ou evitada (*vide* JONSEN; TOULMIN, 1988). A ideia era que as respostas completas a essas perguntas contivessem todas as características das ações, das quais as moralmente relevantes seriam um subconjunto. Embora metafisicamente desinteressante, a ideia de tentar listar todas as características de uma ação dessa maneira representa uma heurística distinta – e extrema – para o raciocínio moral.

Voltando às características moralmente relevantes, uma das teorias mais desenvolvidas é a de Bernard Gert. Ele desenvolve uma lista de características relevantes para saber se a violação de uma regra moral deve ser geralmente permitida. Dada a função projetada da lista, é natural que a maioria de suas características moralmente relevantes faça referência ao conjunto de regras morais defendidas por Gert. Assim, algumas das distinções feitas por Gert entre dimensões de características relevantes refletem posições controversas na teoria moral. Por exemplo, uma das dimensões é se "a violação [é] feita intencionalmente ou apenas conscientemente" (GERT, 1998, p. 234) – uma distinção que aqueles que rejeitam a doutrina do duplo efeito não considerariam relevante.

Ao deliberar sobre o que devemos fazer moralmente, muitas vezes também tentamos descobrir quais considerações são **as mais** relevantes. Usando uma questão mencionada acima: São os contratos de maternidade de aluguel mais parecidos com acordos com babás (claramente aceitáveis) ou com acordos com prostitutas (não tão claramente aceitáveis)? Ou seja, qual característica da maternidade de aluguel é mais relevante: a que envolve um contrato de serviços de cuidadora ou a que envolve o pagamento pelo uso íntimo do corpo? Tanto nesses casos relativamente novos quanto nos mais familiares, o raciocínio por analogia desempenha um papel importante no pensamento moral comum. Quando esse raciocínio por analogia começa a se tornar sistemático – uma conquista social que requer certa estabilidade histórica e reflexão sobre aquilo que é elevado a normas morais – passa a explorar a comparação com casos que são "paradigmáticos", no sentido

de serem tomados como resolvidos. Dentro desta estrutura estável, um sistema de casuística pode desenvolver-se a ponto de estabelecer alguma ordem ao recurso a casos análogos. Para usar uma analogia: a disponibilidade de um conjunto de analogias amplamente aceito e sistemático e a disponibilidade do que são consideradas normas morais podem se comparar como a galinha com o ovo: cada uma pode ser um momento indispensável na gênese do outro.

Assim entendida, a casuística é um auxílio indispensável ao raciocínio moral. Pelo menos, isso se seguiria da junção de duas características da situação moral humana mencionada acima: a multiplicidade de considerações morais que surgem em casos particulares e a necessidade e possibilidade de empregar princípios morais em um raciocínio moral sólido. Exigimos julgamento moral, não simplesmente uma aplicação dedutiva de princípios ou uma intuição particularista sobre o que devemos fazer. Esse julgamento deve responder aos princípios morais, mas não pode ser diretamente derivado deles. Dessa forma, nosso julgamento moral será muito beneficiado se for capaz de se apoiar no tipo de suporte heurístico que a casuística oferece. Pensar em qual dos dois casos análogos fornece uma chave melhor para entender o caso em questão é uma maneira útil de organizar nosso raciocínio moral, da qual devemos continuar a depender. Se não tivermos o tipo de consenso amplo sobre um conjunto de casos paradigmáticos em que os casuístas católicos ou talmúdicos da Renascença podiam se basear, nossos esforços casuísticos serão necessariamente mais controversos e hesitantes do que os deles; mas não estamos totalmente sem casos resolvidos para trabalhar. De fato, como Jonsen e Toulmin sugerem no início de sua extensa explicação e defesa da casuística, a profundidade do desacordo sobre as teorias morais que caracteriza uma sociedade pluralista pode nos levar a ter que atribuir comparativamente **mais** peso aos casos sobre os quais podemos encontrar consenso do que fizeram os casuístas clássicos (JONSEN; TOULMIN, 1988).

Apesar da longa história da casuística, há pouco que possa ser dito de maneira útil sobre como se deve raciocinar sobre analogias concorrentes. Há mais para se estudar no âmbito do Direito, onde os casos anteriores têm importância como precedente. Como observa Sunstein (SUNSTEIN, 1996, cap. 3), O Direito trata de casos particulares, que são sempre “potencialmente distinguíveis” (72); no entanto, a lei também impõe “um requisito de consistência prática” (67). Essa combinação de fatores torna o raciocínio por analogia particularmente influente no Direito, pois é preciso decidir se um determinado caso é mais parecido com um certo conjunto de precedentes ou com outro. Uma vez que a lei deve proceder

mesmo dentro de uma sociedade pluralista como a nossa, argumenta Sunstein, vemos que o raciocínio analógico pode avançar com base em "julgamentos não completamente teorizados" ou no que Rawls chama de "consenso sobreposto" (RAWLS, 1996). Ou seja, embora um uso robusto de casos análogos dependa, como observamos, de algum acordo compartilhado básico, esse acordo não precisa se estender a todos os assuntos ou a todos os níveis do pensamento moral dos indivíduos. Assim, embora em uma sociedade pluralista possamos não ter o tipo de acordo normativo abrangente que tornou possível a alta casuística do cristianismo renascentista, o caminho do Direito sugere que o raciocínio analógico baseado em casos e normativamente influente ainda pode continuar se desenvolvendo. Uma abordagem moderna e concorrente ao raciocínio baseado em casos ou que respeita os precedentes foi desenvolvida por John F. Horty (2016). Na abordagem de Horty, que se baseia na lógica padrão desenvolvida em (HORTY, 2012), o corpo de precedentes sistematicamente altera os pesos das razões que surgem em um novo caso.

Raciocinar por recurso a casos também é o método favorito de alguns filósofos morais recentes. Como nosso foco aqui não está nos métodos da teoria moral, não precisamos entrar em detalhes na comparação de diferentes maneiras pelas quais os filósofos manejam os casos a favor e contra teorias morais alternativas. Há, no entanto, um ponto importante e amplamente aplicável que vale a pena abordar sobre o raciocínio comum com referência a casos, que emerge mais claramente do uso filosófico de tal raciocínio. Os filósofos muitas vezes se sentem livres para imaginar casos, muitas vezes bastante improváveis, a fim de tentar isolar diferenças relevantes. Um exemplo indigesto consiste em dois casos oferecidos por James Rachels para lançar dúvidas sobre o significado moral da distinção entre matar e deixar morrer, ligeiramente reescritos aqui. Em ambos os casos, há no início um menino em uma banheira e um primo mais velho ganancioso no andar de baixo que herdará a mansão da família se e somente se o menino morrer antes dele (RACHELS, 1975). No Caso A, o primo ouve um barulho, corre para o banheiro e ao encontrar o menino inconsciente na banheira estende a mão para abrir a torneira ainda mais para que a água suba e afogue o menino. No caso B, o primo ouve o mesmo barulho, corre para encontrar o menino e ao vê-lo inconsciente na banheira com a água correndo, decide sentar e não fazer nada até que o menino se afogue. Rachels então argumenta que como certamente não há diferença moral entre esses casos, a distinção geral entre matar e deixar morrer é refutada. "Não tão rápido!" é a reação bem justificada (*vide* BEAUCHAMP, 1979). Só porque um fator é moralmente relevante de uma certa maneira na comparação de um par de

casos não significa que ele seja ou deva ser relevante da mesma maneira ou no mesmo grau quando comparado a outros casos. Shelly Kagan nomeou de "falácia aditiva" a falha em levar em conta esse fato de interação contextual na comparação de casos (1988). A partir disso, Kagan conclui que o raciocínio dos teóricos morais deve depender de alguma teoria que nos ajude a antecipar e explicar as maneiras pelas quais os fatores irão interagir em vários contextos. Uma lição paralela, reforçando o que já observamos em conexão com a casuística propriamente dita, aplicar-se-ia ao raciocínio moral em geral: o raciocínio a partir de casos deve, pelo menos, apoiar-se implicitamente em um conjunto de juízos ou crenças organizadoras, de um tipo que, em alguns entendimentos, contam como uma "teoria" moral. Se o ponto estiver correto, ele fornece outro tipo de razão para pensarmos que as considerações morais podem ser cristalizadas em princípios que explicitam a estrutura organizacional envolvida.

2.4. Raciocínio Moral e Psicologia Moral

Estamos ocupados aqui com o raciocínio moral como uma espécie de raciocínio prático – raciocínio direcionado a decidir o que fazer e, se bem-sucedido, gerador de uma intenção. Mas como pode tal raciocínio prático ter sucesso? Como o raciocínio moral pode ligar-se a estados psicológicos motivacionalmente eficazes de modo a ter esse tipo de efeito causal? A "psicologia moral" – nome tradicional para o estudo filosófico da intenção e da ação – tem muito a dizer sobre essas questões, tanto em sua forma tradicional, *a priori,* quanto em sua forma empírica recentemente popular. Além disso, as conclusões da psicologia moral podem ter implicações morais substantivas, pois pode ser razoável supor que, se houver razões profundas para que um determinado tipo de raciocínio moral **não possa** ser colocado em prática, então quaisquer princípios que exijam tal raciocínio não serão sólidos. Nesse espírito, Samuel Scheffler explorou "a importância para a filosofia moral de alguma compreensão razoavelmente realista da psicologia motivacional humana" (SCHEFFLER, 1992, p. 8) e Peter Railton desenvolveu a ideia de que certos princípios morais podem gerar uma espécie de "alienação" (RAILTON, 1984). Em suma, podemos nos interessar naquilo que torna o raciocínio prático de um certo tipo psicologicamente possível tanto pelo valor intrínseco deste problema quanto por seu valor como uma maneira de elaborar parte do conteúdo substancial da teoria moral.

A questão da possibilidade psicológica é importante para todos os tipos de raciocínio prático (*vide* AUDI, 1989). Na moralidade, esta questão é especialmente urgente, pois muitas vezes ela exige que os indivíduos deixem de satisfazer seus próprios interesses. Como resultado, pode parecer que o efeito prático do raciocínio moral não pode ser explicado por um simples apelo às motivações iniciais que moldam ou constituem os interesses de alguém, em combinação com uma obrigação como a mencionada acima, de desejar os meios necessários para seus fins. Ao que parece, a moralidade exige que os indivíduos ajam com objetivos que podem não fazer parte de seu "conjunto motivacional" inicial, na terminologia de Williams (1981). Como o raciocínio moral pode levar as pessoas a fazer isso? A pergunta é antiga. A *República* de Platão respondeu que as aparências enganam e que agir moralmente é, de fato, do próprio interesse do agente esclarecido. Em forte contraste, Kant sustentou que nossa capacidade transcendente de agir de acordo com nossa concepção de uma lei prática nos habilita a estabelecer objetivos e agir moralmente, mesmo quando a ação entra em conflito direto com nossos interesses. Muitas outras propostas têm sido oferecidas. Recentemente, alguns filósofos têm defendido o que chamamos de "internismo motivacional" sobre a moralidade, tese segundo a qual existe uma conexão conceitual necessária entre juízo moral e motivação. Por exemplo, Michael Smith coloca a afirmação da seguinte forma (SMITH, 1994, p. 61):

> Se um agente julga que é correto para ele executar a ação Φ nas circunstâncias C, então ou ele está motivado para executar Φ em C ou ele é agencialmente irracional.

Mesmo essa versão deferida do internismo motivacional pode ser muito forte; mas, em vez de nos aprofundarmos nessa questão, voltemo-nos para uma questão mais interna ao raciocínio moral. (Para mais informações sobre a questão do internismo motivacional, *vide* **motivação moral**.)

Esta tradicional questão surge quando o raciocínio moral é concluído. Supondo que os agentes tenham chegado a uma conclusão moral, pergunta-se como eles podem ser motivados a executá-la. Uma outra questão sobre a intersecção entre raciocínio moral e psicologia moral, mais própria ao primeiro, diz respeito a como os elementos motivacionais moldam o próprio processo de raciocínio.

Uma influente perspectiva filosófica da psicologia humana, derivada de Hume, insiste que crenças e desejos são existências distintas (HUME, 2000, Livro II, parte III, seção III; *vide* SMITH, 1994, p. 7). Isso significa que há sempre um

problema potencial sobre como o raciocínio, que parece funcionar concatenando crenças, liga-se às motivações fornecidas pelo desejo. O elo paradigmático é estabelecido pela instrumentalidade da ação: nas circunstâncias C, o desejo por Ψ liga-se à crença de que a ação Φ levará a conquista de Ψ. Dessa forma, os filósofos que examinam o raciocínio moral dentro de uma psicologia de crença-desejo essencialmente humiana, acabam aceitando uma explicação restrita do raciocínio moral. A própria explicação de Hume exemplifica o tipo de restrição envolvida. Como diz Hume, as paixões calmas sustentam a dupla correção de perspectiva constitutiva da moralidade, aludida acima. Uma vez que essas paixões calmas são vistas como competindo com nossas outras paixões essencialmente no mesmo critério motivacional, por assim dizer, nossas paixões limitam o alcance do raciocínio moral.

Podemos dar um passo significativo no distanciamento de uma compreensão estreita da psicologia moral de Hume se reconhecermos a existência do que Rawls chamou de "desejos derivados de princípios" (RAWLS, 1996, p. 82-83; RAWLS, 2000, p. 46-47). São desejos cujos objetos não podem ser caracterizados sem referência a algum princípio racional ou moral. Um caso especial importante entre estes é o dos "desejos derivados de conceitos", em que o desejo derivado de princípios em questão é visto pelo agente como pertencente a uma concepção mais ampla, e como importante por conta disso (RAWLS, 1996, p. 83-84; 2000, p. 148-152). Por exemplo, concebendo-se como um cidadão, pode-se ter o desejo de contribuir com sua justa quota das necessidades e exigências de uma sociedade. Embora possa parecer que qualquer conteúdo, incluindo este de contribuir com a sociedade, possa substituir Ψ na concepção de desejo de Hume, e embora Hume tenha se proposto a mostrar como sentimentos morais como o orgulho podem ser explicados em termos de mecanismos psicológicos simples, seu influente empirismo na verdade tende a restringir o possível conteúdo dos desejos. Assim, a introdução de desejos derivados de princípios parece marcar um afastamento da psicologia humiana. Como Rawls observa, se "Podemos nos ver atraídos por concepções e ideais expressos tanto pelo [conceito de] correção, quanto pelo [conceito de] bem... como fixar limites para o que, no pensamento e na deliberação, pode mover as pessoas, e, consequentemente, a partir dos quais as pessoas podem agir? (1996, p. 85; trad. Bras., p. 130). Enquanto Rawls desenvolveu este ponto contrastando a psicologia moral de Hume com a de Kant, o mesmo ponto básico também é feito por neoaristotélicos (*vide* MCDOWELL, 1998).

A introdução de desejos derivados de princípios rompe qualquer limite naturalista ao seu conteúdo; no entanto, alguns filósofos sustentam que essa noção

permanece muito apegada a uma tradição essencialmente humiana para ser capaz de capturar a ideia de comprometimento moral. Ela mantem em aberto o entendimento de que os desejos se mantem como os itens motivacionais que competem entre si com base em sua força. Dizer que o desejo de ser justo pode ser superado pelo desejo de progresso individual pode falhar em capturar o pensamento de que se tem um compromisso moral com a justiça – mesmo que não absoluto. Sartre elaborou seu exemplo do estudante dividido entre ficar com sua mãe e lutar com a França Livre de modo a fazer parecer implausível que ele devesse decidir simplesmente determinando o que ele mais fortemente deseja fazer.

Uma maneira de chegar à ideia de compromisso moral é enfatizar nossa capacidade de refletir sobre o que queremos. Por esse caminho, pode-se distinguir, à moda de Harry Frankfurt, entre a força de nossos desejos e "a importância daquilo que nos importa" (FRANKFURT, 1988). Embora essa ideia seja sedutora, ela oferece uma visão relativamente rasa sobre **como** é que refletimos. Outra maneira de modelar o comprometimento moral é assumir que nossas intenções operam em um nível distinto de nossos desejos, estruturando o que estamos dispostos a reconsiderar em qualquer ponto de nossas deliberações (*vide* BRATMAN, 1999). Embora essa abordagem em duplo nível ofereça algumas vantagens, ela é limitada pela concessão de uma espécie de primazia normativa, no nível irrefletido, aos desejos não-derivados. Uma abordagem mais integrada pode modelar a psicologia do compromisso moral de forma a redesenhar a natureza do desejo desde as bases da teoria. Outra possibilidade atraente é retornar à concepção aristotélica do desejo, em que ele existe em função de algum bem ou bem aparente (*vide* RICHARDSON, 2004). Nessa concepção, o fim pelo qual uma ação é executada desempenha um importante papel regulador, indicando, pelo menos em parte, o que **não** se deve fazer (RICHARDSON, 2018, §§ 8.3-8.4). O raciocínio sobre nossos fins últimos tem, portanto, um caráter distintivo (*vide* RICHARDSON, 1994; SCHMIDTZ, 1995). Qualquer que seja a melhor explicação filosófica da noção de comprometimento moral, muito do nosso raciocínio moral parece envolver expressões desses compromissos e ao mesmo tempo vetores contrários a esses compromissos (ANDERSON; PILDES, 2000). Para uma alternativa, confira o trabalho de Tiberius (2000).

Trabalhos empíricos recentes, empregando tecnologias de pesquisa e de imagem cerebral, tem permitido aos filósofos abordar questões sobre a base psicológica do raciocínio moral a partir de novos ângulos. Os resultados iniciais das pesquisas sobre o cérebro parecem mostrar que indivíduos com danos no lobo pré-frontal tendem a raciocinar de maneira mais diretamente consequencialista do que

aqueles sem tais danos (KOENIGS *et al*., 2007). Alguns teóricos consideram que esta descoberta tende a confirmar que o raciocínio moral humano totalmente competente vai além de uma simples pesagem de prós e contras para incluir a avaliação de restrições morais (*vide* WELLMAN; MILLER, 2008; YOUNG; SAXE, 2008). No entanto, outros argumentam que as respostas emocionais do lobo pré-frontal interferem no raciocínio mais sóbrio, sólido e de estilo consequencialista das outras partes do cérebro (*vide* GREENE, 2014). Os dados da pesquisa revelam ou confirmam, entre outras coisas, interessantes assimetrias normativamente carregadas em nosso emprego de conceitos como responsabilidade e causalidade (KNOBE, 2006). Também revela que muitas das distinções mais sutis da teoria moral, como a distinção entre um meio pretendido e um efeito colateral previsto, estão profundamente embutidas em nossas psicologias, estando presentes nas mais diversas culturas e em crianças pequenas, de uma forma que leva alguns a pensar na possibilidade de uma "**gramática moral**" inata (MIKHAIL, 2011).

Uma última questão sobre a conexão entre motivação moral e raciocínio moral recai sobre se alguém desprovido dos compromissos motivacionais adequados poderia raciocinar corretamente dentro do domínio moral. A resposta seria positiva na concepção oficial e estrita de Hume, onde o ato de raciocinar é essencialmente restrito a traçar conexões empíricas e lógicas. O agente imoral poderia traçar as mesmas implicações causais e lógicas de uma ação que uma pessoa moralmente virtuosa. A única diferença seria prática, não racional: eles não agirão da mesma forma. Observe, no entanto, que a resposta humiana afirmativa depende do afastamento da definição de "raciocínio moral" usada neste artigo, que o apresenta como uma espécie de raciocínio prático. Notavelmente, Kant pode responder "sim" enquanto ainda apresenta o raciocínio moral como prático. De acordo com Kant, na *Fundamentação* e na *Crítica da Razão Prática*, raciocinar bem, no âmbito moral, não depende de nenhum comprometimento motivacional prévio, mas permanece como raciocínio prático. Isso porque ele acredita que a própria lei moral pode gerar motivação em si mesma. (Kant oferece uma psicologia mais complexa em *A Metafísica dos Costumes* e *Religião*.) Para Aristóteles, em contraste, um agente cujas motivações não são constituídas de forma virtuosa irá sistematicamente se enganar sobre o que é bom e o que é ruim e, portanto, será incapaz de raciocinar com excelência. O melhor raciocínio de que uma pessoa viciosa é capaz, segundo Aristóteles, é um simulacro defeituoso de sabedoria prática que ele chama de "astúcia" (*Ética a Nicômaco* 1144a25; Trad. Bras., 1991, p.137).

2.5. Modelagem de Considerações Morais Conflitantes

Frequentemente, as considerações morais entram em conflito umas com as outras. E o mesmo ocorre com os princípios morais e os comprometimentos morais. Assumindo que lealdade filial e patriotismo são considerações morais, então o aluno de Sartre enfrenta um conflito moral. Lembremo-nos de que uma coisa é modelar a metafísica da moralidade ou as condições de verdade das declarações morais e outra é oferecer uma teoria do raciocínio moral. Olhando agora para considerações conflitantes, nosso interesse aqui permanece na teorização do raciocínio moral e não em conflitos internos às teorias morais de primeira ordem. Nosso principal interesse está nas maneiras pelas quais precisamos estruturar ou pensar sobre considerações conflitantes para bem conduzir nosso raciocínio envolvendo-as.

Um instrumento influente para se pensar sobre conflitos morais é a noção de um "dever *prima facie*", de W. D. Ross. O termo se tornou popular apesar de enganosamente sugerir mera aparência – a forma como as coisas parecem à primeira vista. Alguns filósofos morais preferem o termo "dever *pro tanto*" (por exemplo, HURLEY, 1989). Ross explicou que seu termo fornece "uma breve maneira de se referir à característica (bastante distinta daquela de ser um dever propriamente dito) que um ato tem, em virtude de ser de um certo tipo (por exemplo, o cumprimento de uma promessa), de ser um ato que seria realmente um dever se não fosse ao mesmo tempo de outro tipo moralmente significativo". Ilustrando o ponto, ele observou que um dever *prima facie* de manter uma promessa pode ser superado por um dever *prima facie* de evitar um acidente grave, resultando em um dever genuíno ou incondicional de evitar o acidente (ROSS, 1988, p. 18-19). Ross descreveu cada dever *prima facie* como um atributo "parti-resultante", fundamentado ou explicado por um aspecto de um ato, enquanto "ser um dever [real]" é um atributo "toti-resultante" derivado de todos esses aspectos de um ato, tomados em conjunto (ROSS, 28; *vide* PIETROSKI, 1993). Isso sugere que em cada caso há, em princípio, alguma função que geralmente converte as contribuições parciais de cada dever *prima facie* em algum dever real. Qual seria essa função? Para crédito de Ross, ele escreve que "até onde eu posso ver, nenhuma regra geral pode ser estabelecida para o cálculo da rigorosidade normativa relativa dessas obrigações *prima facie*" (41). Assim, uma segunda vertente em Ross simplesmente enfatiza, seguindo Aristóteles, a necessidade de julgamento prático por aqueles que são educados a ser virtuosos (42).

Como podem as considerações do tipo constituído por deveres *prima facie* entrar em nosso raciocínio moral? Elas podem fazê-lo explicitamente, ou apenas

implicitamente. Há também uma terceira possibilidade, ainda mais tênue (SCHEFFLER, 1992, p. 32): pode ser simplesmente o caso de que, se o agente **tivesse** reconhecido um dever *prima facie*, ele teria agido de acordo com este dever, a menos que o considerasse superado por outro dever. E isto é um fato sobre como ele **teria** raciocinado.

Apesar de Ross negar que exista qualquer método geral para calcularmos a adstringência normativa relativa dos deveres *prima facie*, há uma outra vertente em sua exposição que muitos acham irresistível e que tende a minar essa negação. No mesmo parágrafo em que afirma que não vê regras gerais para lidar com conflitos, ele fala em termos de "o maior equilíbrio do que é *prima facie* correto". Essa linguagem, juntamente com a ideia de "rigorosidade normativa relativa", inelutavelmente sugere a ideia de que a função de conversão pode ser a mesma em cada caso de conflito e que pode ser quantitativa. Nesta concepção, se houver um conflito entre dois deveres *prima facie*, aquele que for mais forte naquelas circunstâncias deve ser considerado vencedor. Devidamente advertidos sobre a falácia aditiva (ver seção 2.3)**,** podemos reconhecer que a força de uma consideração moral em um conjunto de circunstâncias não pode ser inferida a partir de sua força em outras circunstâncias. Portanto, em muitos casos essa abordagem ainda precisará contar com julgamentos intuitivos. Mas esse julgamento intuitivo será sobre qual consideração *prima facie* é mais forte nas circunstâncias dadas, não simplesmente sobre o que deve ser feito.

O pensamento de que nosso raciocínio moral requer ou é beneficiado por um improviso quantitativo desse tipo tem uma longa linhagem. Podemos realmente raciocinar bem moralmente de uma maneira que se reduza a avaliar os pesos das considerações concorrentes? Abordar essa questão exigirá um excurso sobre a natureza das razões morais. O suporte filosófico para essa possibilidade envolve uma ideia de comensurabilidade prática. Precisamos distinguir, aqui, dois tipos de comensurabilidade ou incomensurabilidade prática, uma definida em termos metafísicos e outra em termos deliberativos. Cada uma dessas formas pode ser declarada avaliativamente ou deonticamente. O tipo metafísico de incomensurabilidade entre valores é definido diretamente em termos do que é o caso. Assim, para apresentar uma versão avaliativa: dois valores são metafisicamente incomensuráveis apenas no caso de nenhum ser melhor que o outro nem serem igualmente bons (ver CHANG, 1998). Agora, a incomensurabilidade metafísica dos valores, ou sua ausência, está apenas vagamente ligada a como seria deliberar de maneira razoável. Se todos os valores ou considerações morais são metafisicamente (ou seja, de fato) comensuráveis, ainda assim pode ser que nosso acesso à função calculadora última seja tão limitado que nos sairíamos mal ao prosseguir em nossas deliberações

para tentar decidir sobre quais resultados são “melhores” ou quais considerações são “mais fortes”. Podemos não ter ideia de como medir a “força” relevante. Inversamente, mesmo que a incomensurabilidade metafísica dos valores seja comum, podemos ter algum sucesso deliberativo ao proceder como se não fosse esse o caso, assim como procedemos na termodinâmica como se as leis dos gases fossem obtidas em sua forma idealizada. Assim, ao pensar nas implicações deliberativas de valores incomensuráveis, faríamos bem em pensar em termos de uma definição adaptada ao contexto deliberativo. Comece com uma formulação local, por pares. Podemos dizer que duas opções, A e B, são deliberativamente comensuráveis apenas no caso de haver alguma dimensão de valor em termos da qual é possível representar adequadamente a força das considerações relevantes para a escolha, e onde a representação está disponível antes da escolha ou de forma logicamente independente dela.

Nesse sentido, filósofos tão diversos quanto Immanuel Kant e John Stuart Mill argumentaram que, a menos que duas opções sejam deliberativamente comensuráveis, é impossível escolher racionalmente entre elas. Curiosamente, Kant limitou essa afirmação ao domínio das considerações prudenciais, reconhecendo o raciocínio moral como invocando considerações incomensuráveis com as considerações prudenciais. Para Mill, essa tese formou uma parte importante de seu argumento de que deve haver algum princípio que atue como “árbitro” final – ou seja, de acordo com ele, o princípio da utilidade. Henry Sidgwick elaborou o argumento de Mill e tornou explícito seu pressuposto crucial, que ele chamou de “princípio da validade superior” (SIDGWICK, 1981; SCHNEEWIND, 1977). De acordo com este princípio, o conflito entre considerações morais ou práticas distintas pode ser racionalmente resolvido apenas com base em algum terceiro princípio ou consideração que seja mais geral e mais firmemente garantido do que os dois concorrentes iniciais. A partir dessa suposição, pode-se prontamente construir um argumento, como o fizeram Mill e Sidgwick, para a necessidade racional não apenas de comensurabilidade deliberativa local, mas de uma comensurabilidade deliberativa global que aceita apenas um princípio como árbitro final (RICHARDSON, 1994, cap. 6).

A clareza de Sidgwick é valiosa também para nos ajudar a ver como resistir à demanda por comensurabilidade deliberativa. A comensurabilidade deliberativa não é necessária para proceder racionalmente se as considerações conflitantes puderem ser racionalmente tratadas de uma forma holística que não envolva o apelo a um princípio de “validade superior”. Que nosso raciocínio moral pode proceder de forma holística é fortemente defendido por Rawls. As caracterizações

de Rawls do influente ideal de equilíbrio reflexivo e suas ideias relacionadas sobre a natureza da justificação implicam que podemos lidar com considerações conflitantes de maneiras menos hierárquicas do que imaginado por Mill ou Sidgwick. Em vez de proceder como subindo uma escada de apelação até algum tribunal superior ou árbitro supremo, Rawls sugere que "trabalhemos partindo de ambos os lados" quando enfrentamos considerações conflitantes (RAWLS, 1999, p. 18). Às vezes, de fato, revisamos nossos julgamentos mais particulares à luz de algum princípio geral ao qual aderimos; mas também somos livres para revisar princípios mais gerais à luz de algum julgamento relativamente concreto que consideramos. Neste quadro, não há correlação necessária entre grau de generalidade e força de autoridade normativa. Que esta maneira holística de proceder (seja na construção da teoria moral ou na deliberação: *vide* HURLEY, 1989) pode ser racional é confirmado pela possibilidade de uma forma de justificação que é igualmente holística: "a justificação é uma questão de apoio mútuo entre muitas considerações, de tudo se encaixando em uma visão coerente" (RAWLS, 1999, p. 19; p. 507). (Note-se que esta tese, que expressa um aspecto necessário da justificação moral ou prática, não deve ser tomada como uma definição ou análise das justificações morais ou práticas). Portanto, há uma alternativa à dependência de encontrar uma dimensão deliberativa em termos da qual as considerações possam ser classificadas como "mais forte" ou "melhor" ou "mais adstringente": pode-se, em vez disso, "moldar e ajustar" [itens dentro do conjunto holístico] com o objetivo de construir maior apoio mútuo entre as considerações que o agente endossa durante devida reflexão. Se até mesmo o resultado desejado da coerência prática está sujeito a tal re-especificação, então essa possibilidade holística realmente representa uma alternativa à comensurabilidade, já que o deliberador, e não algum padrão de coerência, retém a soberania reflexiva (RICHARDSON, 1994, sec. 26). O resultado pode ser aquele em que as considerações originalmente concorrentes não são tanto comparadas, mas transformadas (RICHARDSON, 2018, cap. 1).

Suponha que comecemos com um conjunto de considerações morais de primeira ordem que são todas comensuráveis como uma questão de fato metafísica última, mas que nossa compreensão da força real dessas considerações é bastante pobre e sujeita a distorções sistemáticas. Talvez algumas pessoas estejam melhor posicionadas do que outras para apreciar certas considerações, e talvez nossas interações estratégicas nos levem a alcançar resultados abaixo do ideal se cada um de nós perseguir seu próprio julgamento livre acerca de como o conjunto geral de considerações se desenrola. Em tais circunstâncias, há um forte argumento para

abandonarmos o raciocínio por maximização , sem com isto mergulharmos inteiramente na alternativa holística. Este caso foi articulado de forma influente por Joseph Raz, que desenvolve a noção de uma "razão excludente" para ocupar essa posição intermediária (RAZ, 1990).

"Uma razão excludente", na terminologia de Raz, "é uma razão de segunda ordem para se abster de agir por alguma razão" (p. 39). Um exemplo simples é o de Ann, que está cansada após um dia longo e estressante e, portanto, tem razões para não agir de acordo com sua melhor avaliação das razões relacionadas a uma decisão de investimento particularmente importante que ela enfrenta no momento (p. 37). Essa noção de razão excludente permitiu a Raz capturar muitas das complexidades do nosso raciocínio moral, especialmente porque envolve comprometimentos baseados em princípios, embora admitindo que, em primeira ordem, todas as razões práticas podem ser comensuráveis. A estratégia inicial de Raz para conciliar a comensurabilidade com a complexidade da estrutura foi limitar a tese de que as razões são comparáveis, com respeito à força, a razões de uma outra ordem. Razões de primeira ordem competem com base na força; mas os conflitos entre razões de primeira e segunda ordem "são resolvidos não pela força das razões concorrentes, mas por um princípio geral de raciocínio prático que determina que as razões excludentes sempre prevalecem" (p. 40).

Se aceitarmos o "princípio geral do raciocínio prático", por que deveríamos reconhecer a existência de quaisquer razões excludentes, que por definição prevalecem independentemente de qualquer disputa de força? A principal resposta de Raz a essa pergunta abandona o domínio metafísico das forças que várias razões "possuem" para abordar o ponto de vista epistemicamente limitado do deliberador. Como no caso de Ann, podemos ver que em certos contextos um deliberador provavelmente cometerá erros se agir de acordo com sua percepção das razões de primeira ordem. Razões de segunda ordem indicam, com respeito a uma certa gama de razões de primeira ordem, que o agente "não deve agir **por** essas razões" (185). A justificativa mais ampla para a existência de razões excludentes pode ser então consistentemente colocada em termos de razões de primeira ordem comensuráveis. Essa justificativa pode assumir a seguinte forma: "Dadas as limitações deliberativas deste agente, o equilíbrio das razões de primeira ordem provavelmente será melhor conformado se o agente se abster de agir por algumas dessas razões".

A explicação de Raz das razões excludentes pode ser usada para reconciliar a comensurabilidade última com a complexidade estruturada presentes no nosso raciocínio moral. O êxito de tal tentativa dependeria, em parte, da extensão em que

temos uma compreensão real das razões de primeira ordem, cujos conflitos podem ser resolvidos apenas com base em sua força comparativa. Nossas considerações, acima, sobre a casuística, a falácia aditiva e a incomensurabilidade deliberativa podem se combinar para fazer parecer que apenas em raras ocasiões de nossa prática temos uma boa compreensão das razões de primeira ordem, se estas forem definidas, à la Raz, como competindo apenas em termos de força. Se isto estiver correto, quase sempre teremos boas razões excludentes para raciocinar em alguma outra base que não em termos da força relativa das razões de primeira ordem. Sob essas suposições, a terceira via que a ideia de razões excludentes de Raz parece desbravar se aproximaria mais do método holístico.

A noção de "força" de uma consideração moral, seja apresentada como parte de um quadro metafísico de como as considerações de primeira ordem interagem de fato ou como uma sugestão sobre como resolver um conflito moral, não deve ser confundida com a determinação última sobre se uma consideração, e especificamente um dever, se sobrepõe a outro. No exemplo de Ross sobre deveres *prima facie* conflitantes, o agente deve escolher entre evitar um acidente grave e manter a promessa de encontrar com um amigo. (Ross escolheu o caso para ilustrar que um dever "imperfeito", ou um dever de comissivo, pode se sobrepor a um dever estrito e proibitivo). A suposição de Ross é que, neste caso, todas as pessoas bem-educadas concordariam que o dever de evitar danos graves a alguém se sobrepõe ao dever de cumprir aquela promessa. Podemos considerar, se quisermos, que esse julgamento implica que, neste caso, consideramos o dever de salvar uma vida mais forte do que o dever de cumprir uma promessa; mas, na verdade, essa tese sobre a força relativa das considerações não acrescenta nada à nossa compreensão da situação. Pois, não chegamos à nossa conclusão prática neste caso determinando que o dever de salvar a vida do menino é mais forte. A afirmação de que este dever é mais forte é simplesmente uma maneira de embelezar a conclusão de que, entre os dois deveres *prima facie* aqui em conflito, o mais forte é aquele que expressa o dever *omnibus perpensis*. Ser "sobreposto" é apenas ser um dever *prima facie* que não gera um dever efetivo porque outro dever *prima facie* que entra em conflito com ele – ou vários deles que o fazem – gera um dever efetivo. Assim, o julgamento de que alguns deveres se sobrepõem a outros pode ser entendido apenas em termos de seus resultados deônticos e sem referência a considerações de força. Para confirmar isso, observe que podemos dizer: "Por uma questão de fidelidade, devemos manter a promessa; por uma questão de beneficência, devemos salvar uma vida; não podemos fazer as duas coisas; e ambas as categorias consideram que deveríamos salvar uma vida."

Compreender a noção de um dever sobrepondo-se a outro dessa maneira nos coloca em posição de abordar o tópico dos dilemas morais. Como esse tópico é abordado em um artigo separado, aqui podemos simplesmente adotar uma definição atraente de dilema moral. Sinnott-Armstrong (1988) sugeriu que um dilema moral é uma situação em que o seguinte é verdadeiro para um único agente:

1 Ele deve executar a ação A
2 Ele deve executar a ação B
3 Ele não pode executar ambas A e B
4 (1) não se sobrepõe a (2) e (2) não se sobrepõe a (1)

Essa maneira de definir dilemas morais os distingue do tipo de conflito moral, como o dilema de Ross entre o cumprimento de promessas versus a prevenção de acidentes, em que um dos deveres é superado pelo outro. Alegadamente, o aluno de Sartre enfrenta um dilema moral. Dar sentido a uma situação em que nenhum dos dois deveres se sobrepõe ao outro é mais fácil se a comensurabilidade deliberativa for negada. Se os dilemas morais são possíveis dependerá crucialmente do debate sobre se "dever" implica "poder" e se qualquer par de deveres como aqueles compreendidos por (1) e (2) implica um único dever "aglomerado" que o agente faça tanto A quanto B. Se qualquer um desses supostos princípios da lógica deôntica for falso, então dilemas morais são possíveis.

Jonathan Dancy destacou bem um tipo de variabilidade contextual nas razões morais que veio a ser conhecido como "holismo de razões": "uma característica [de uma ação] pode constituir uma razão em um caso, assim como pode, em outros casos, não ser razão alguma, ou até mesmo uma razão oposta" (DANCY, 2004). Para adaptar um de seus exemplos: embora muitas vezes haja uma razão moral para não mentir, quando se joga o pôquer mentiroso, geralmente deve-se mentir; caso contrário, estraga-se o jogo (*vide* DANCY, 1993, p. 61). Dancy argumenta que o holismo de razões justifica o particularismo moral do tipo discutido na seção 2.2, segundo o qual não há princípios morais defensáveis. Levar essa conclusão a sério afetaria radicalmente a forma como conduzimos nosso raciocínio moral. A premissa holista do argumento foi desafiada (por exemplo, AUDI, 2004, MCKEEVER, & RIDGE, 2006). Alguns filósofos também desafiaram de várias maneiras a inferência do holismo de razões para o particularismo. Mark Lance e Margaret Olivia Little (2007) fizeram isso ao mostrar como generalizações revogáveis, na ética e em outras áreas, dependem **sistematicamente** de seu contexto. Eles sugerem que

podemos trabalhar com as generalizações utilizando uma habilidade semelhante à capacidade de discernir considerações moralmente salientes, ou seja, a habilidade de discernir semelhanças relevantes entre mundos possíveis. De forma mais geral, John F. Horty desenvolveu uma explicação lógica e semântica segundo a qual as razões são revogáveis e, portanto, se comportam de forma holística, mas ainda assim existem princípios gerais que explicam como elas se comportam (HORTY, 2012). Por fim, Mark Schroeder argumentou que nossas visões holísticas sobre razões são, na verdade, melhor explicadas supondo que existam princípios gerais (SCHROEDER, 2011).

Este excurso sobre as razões morais sugere que há uma série de boas razões pelas quais o raciocínio sobre questões morais pode não se reduzir simplesmente a avaliar os pesos de considerações concorrentes.

2.6. Aprendizagem Moral e a Revisão das Teorias Morais

Se temos algum conhecimento moral, seja sobre princípios morais gerais ou conclusões morais concretas, certamente é muito imperfeito. O conhecimento moral de que somos capazes dependerá, em parte, de que tipo de raciocínio moral somos capazes. Embora algum aprendizado moral possa resultar do trabalho teórico de filósofos e teóricos morais, muito do que aprendemos com relação à moralidade certamente surge no contexto prático da deliberação sobre casos novos e espinhosos. Essa deliberação pode ser meramente instrumental, ocupada apenas em estabelecer meios para fins morais, ou em estabelecer esses fins. Não há problema especial em aprender o que conduz a fins moralmente obrigatórios: isso é uma questão comum de aprendizado empírico. Mas por meio de que tipo de processo podemos aprender quais fins são moralmente obrigatórios ou quais normas são moralmente exigidas? E, mais especificamente, o aprendizado estritamente moral é possível por meio do raciocínio moral?

Muito do que foi dito acima em relação à abordagem moral se aplica novamente neste contexto, com aproximadamente o mesmo grau de dubiedade ou persuasão. Se há uma função para a percepção moral ou para as emoções na tomada de consciência das considerações morais pelos agentes, elas também podem funcionar para guiar os agentes a novas conclusões. Por exemplo, é concebível que nossa capacidade de indignação seja um detector relativamente confiável de ações erradas, mesmo inéditas, ou que nossa capacidade de prazer

seja um detector confiável de ações que valem a pena executar, mesmo inéditas. Para uma defesa completa desta última possibilidade, que intrigantemente interpreta o prazer como um juízo de valor, confira Millgram (1997). Talvez essas capacidades de julgamento emocional permitam o aprendizado estritamente moral da mesma forma que as sensibilidades treinadas dos jogadores de xadrez os permitem reconhecer a ameaça em uma situação inédita no tabuleiro de xadrez (LANCE; TANESINI, 2004). Ou seja, talvez nossas emoções morais desempenhem um papel crucial no exercício de uma habilidade pela qual nos tornamos capazes de articular internalizações morais que nunca antes alcançamos. Talvez as considerações morais concorrentes interajam de maneiras contextualmente específicas e complexas, tanto quanto as considerações concorrentes do xadrez. Se assim for, faria sentido confiar em nossas capacidades de julgamento emocionalmente guiadas para lidar com complexidades que não podemos modelar explicitamente, mas também esperar que, uma vez guiados, possamos, em retrospectiva, articular algo sobre a lição de uma situação bem navegada.

Um modelo diferente de aprendizagem estritamente moral coloca a ênfase em nossas reações após o fato, em vez de qualquer orientação prévia emocional ou crítica: o modelo de "experimentos vivenciais", para usar a expressão de John Stuart Mill (*vide* ANDERSON, 1991). Aqui, o pensamento básico é que podemos tentar algo e ver se "funciona". Para que isso seja uma alternativa ao aprendizado empírico sobre o que causalmente conduz a quê, precisamos permanecer abertos quanto ao que queremos dizer com as coisas "funcionando". Na terminologia de Mill, por exemplo, precisamos permanecer abertos sobre quais são as "partes" importantes da felicidade. Se assim fizermos, então talvez possamos aprender por experiência quais são algumas dessas partes – isto é, quais são alguns dos meios constitutivos da felicidade. Esses pensamentos unidos, de que nossa vida prática é experimental e que não temos uma concepção firmemente fixa do que significa algo "funcionar", vêm à tona na ética pragmatista de Dewey (ver especialmente DEWEY, 1967 [1922]). Essa concepção experimentalista de aprendizagem estritamente moral é aplicada ao raciocínio moral nas eloquentes caracterizações de Dewey da "inteligência prática" como envolvendo uma abordagem criativa e flexível para descobrir "o que funciona" de uma maneira completamente aberta a repensar nossos objetivos finais.

Uma vez que reconheçamos que o aprendizado moral é uma possibilidade para nós, podemos reconhecer uma gama mais ampla de maneiras de lidar com conflitos morais do que as examinadas na seção anterior. Lá, os conflitos morais

eram descritos de uma forma que pressupunha que o conjunto de considerações morais, entre as quais os conflitos surgiam, deveria ser tomado como fixo. No entanto, se podemos aprender moralmente, então provavelmente podemos e devemos revisar o conjunto de considerações morais que reconhecemos. Muitas vezes, fazemos isso reinterpretando algum princípio moral com o qual começamos, seja tornando-o mais específico, tornando-o mais abstrato ou de alguma outra forma (*vide* RICHARDSON, 2000, 2018).

2.7. Como Podemos Raciocinar Moralmente Uns Com Os Outros?

Até agora, discutimos o raciocínio moral principalmente como se fosse um empreendimento solitário. Esta é, na melhor das hipóteses, uma simplificação conveniente. Na pior das hipóteses, como Jürgen Habermas há muito argumenta, ela distorce profundamente o caráter essencialmente dialógico ou conversacional do raciocínio (*vide* HABERMAS, 1984; LADEN, 2012). De qualquer forma, é claro que muitas vezes precisamos raciocinar moralmente uns com os outros.

Aqui, estamos interessados em como as pessoas podem realmente raciocinar umas com as outras – não no que se pode aprender com como raciocinam coletivamente participantes imaginários em uma posição original ou situação ideal de fala, o que é uma preocupação para a teorização moral propriamente dita. Há duas maneiras salientes e distintas de pensar sobre as pessoas que raciocinam moralmente umas com as outras: 1. como membros de um corpo organizado ou corporativo que é capaz de tomar suas próprias decisões práticas; e 2. como indivíduos autônomos trabalhando fora de tal estrutura para descobrir uns com os outros o que deveriam fazer moralmente.

A natureza e a possibilidade de raciocínio coletivo dentro de um corpo coletivo organizado tem sido recentemente objeto de alguma discussão. Os coletivos podem raciocinar se estiverem estruturados como um agente. Essa estrutura pode ou não ser institucionalizada. Em consonância com a exposição do raciocínio oferecida acima, que pressupõe que um agente é guiado por uma avaliação das próprias razões, é plausível sustentar que um agente coletivo "conta como raciocinando, não apenas racional, apenas se for capaz de formar não apenas crenças ***em*** proposições – isto é, crenças na linguagem objeto – mas também crenças ***sobre*** proposições" (LIST e PETTIT, 2011, 63). Como List e Pettit mostraram (2011, 109-113), os participantes de um agente coletivo inevitavelmente terão incentivos para

deturpar suas próprias preferências em condições que envolvem desacordos ideologicamente estruturados, onde as partes em conflito são orientadas a alcançar ou evitar certos resultados – como às vezes é o caso em que surgem sérias divergências morais. Em contextos em que, em última análise, o que importa é quão bem o grupo ou coletivo relevante acaba se saindo, o "raciocínio de equipe" orientado para o florescimento coletivo do grupo pode ajudá-lo a alcançar um resultado coletivamente ótimo (SUGDEN, 1993, BACHARACH, 2006; ver artigo intencionalidade coletiva). Entretanto, quando o grupo em questão é menor do que o conjunto de pessoas, este foco coletivamente prudencial é distinto de um foco moral e parece estar em desacordo com o tipo de imparcialidade tipicamente considerado distintivo do ponto de vista moral. Pensar em como seria a forma de uma "orientação de equipe" para o conjunto de todas as pessoas pode nos trazer de volta aos pensamentos da universalização kantiana; mas lembremo-nos de que aqui estamos focados no raciocínio real, não no raciocínio hipotético. No que diz respeito ao raciocínio real, mesmo que os indivíduos possam assumir esta orientação para a "equipe" de todas as pessoas, há uma razão séria, destacada por outra vertente da tradição kantiana, para duvidar que qualquer indivíduo possa abandonar seu juízo moral em função do veredicto de qualquer grupo (WOLFF, 1998).

Isso não significa que as pessoas não possam raciocinar juntas, moralmente. Sugere, no entanto, que tal raciocínio conjunto é melhor perseguido como uma questão onde agentes morais independentes trabalham juntos o que devem fazer em relação a uma questão sobre a qual eles têm alguma necessidade de cooperar. Mesmo se for impossível que um agente endosse o veredito de outro agente sobre como se deve agir moralmente, ainda é possível que se possa licitamente levar em conta o testemunho moral de outros. Para opiniões divergentes, confira Mcgrath (2009) e Enoch (2014).

No caso de indivíduos independentes raciocinando moralmente uns com os outros, podemos entender o desacordo moral como uma oportunidade e não um obstáculo. Certamente, se o desacordo moral dos indivíduos for muito profundo, eles podem não conseguir desenvolver esse raciocínio; mas, como nos lembra o exemplo de Kant sobre Carlos V e seu irmão, ambos ansiando por Milão, desacordos intratáveis também podem surgir de desacordos que, embora conceitualmente superficiais, são circunstancialmente agudos. Se for verdade que uma justificação lúcida das crenças morais exige que o agente as veja como sendo, em última análise, fundamentadas em princípios a priori, como G.A. Cohen argumentou (COHEN, 2008, cap. 6), então o espaço para os indivíduos resolverem seus

desacordos morais raciocinando uns com os outros será relativamente restrito; mas é seriamente duvidoso que a natureza da (lúcida) fundamentação moral é realmente tão restrita (RICHARDSON, 2018, § 9.2). Em contraste com o que essa tese sugere, os compromissos morais dos indivíduos parecem suficientemente abertos para serem repensados e as pessoas parecem capazes de se engajar em compromissos baseados em princípios – ou seja, não agem simplesmente visando minimizar perdas (RICHARDSON, 2018, §8.5).

E quanto à possibilidade de que a comunidade moral, aproximadamente a comunidade de todas as pessoas, possa raciocinar como um todo? Essa possibilidade não levanta o tipo de ameaça à imparcialidade que é levantada pelo raciocínio de equipe de um grupo menor de pessoas; mas é difícil vê-lo funcionando de uma maneira que não entre em conflito com a preocupação sobre se qualquer pessoa pode endossar adequadamente, em um sentido forte, os juízos morais de outro agente. Mesmo assim, resta uma possibilidade residual, que é a de que a comunidade moral possa raciocinar de uma única maneira, ou seja, aceitando ou ratificando uma conclusão moral que já se tornou compartilhada de maneira suficientemente abrangente e inclusive (RICHARDSON, 2018, cap. 7).

Referência bibliográfica

ANDERSON, E. S., John Stuart Mill and experiments in living, **Ethics**, 102: 4–26, 1991.

ANDERSON, E. S. and PILDES, R. H., Expressive theories of law: A general restatement**, University of Pennsylvania Law Review**, 148: 1503–1575; 2000.

ARPALY, N. and SCHROEDER, T. **In praise of desire**, Oxford: Oxford University Press.

AUDI, R**., Practical reasoning**, London: Routledge, 1989.

AUDI, R**., The good in the right: A theory of good and intrinsic value**, Princeton: Princeton University Press, 2004.

BACHARACH, M., **Beyond individual choice: Teams and frames in game theory**, Princeton: Princeton University Press, 2006

BEAUCHAMP, T. L., A reply to Rachels on active and passive euthanasia, in ed., CLIFTON W. L. R, **Medical responsibility**, N.J.: Humana Press, 182–95, 1979.

BRANDT, R. B., **A theory of the good and the right**, Oxford: Oxford University Press, 1979.

BRATMAN, M., **Faces of intention: Selected essays on intention and agency**, Cambridge, England: Cambridge University Press, 1999.

BROOME, J., The unity of reasoning? in ed. ROBERTSON S. **Spheres of reason**, Oxford: Oxford University Press, 2009.

BROOME, J., **Rationality through Reasoning**, Chichester, West Sussex: Wiley Blackwell, 2013.

CAMPBELL, R. and KUMAR, V., Moral reasoning on the ground, **Ethics**, 122: 273–312, 2012.

CHANG, R. (ed.), **Incommensurability, incomparability, and practical reason**, Cambridge, Mass.: Harvard University Press, 1998.

CLARKE, S. G., and SIMPSON, E., **Anti-theory in ethics and moral conservativism**, Albany: SUNY Press, 1989.

DANCY, J., **Moral reasons**, Oxford: Blackwell, 1993.

DANCY, J., **Ethics without principles**, Oxford: Oxford University Press, 2004.

DEWEY, J., The middle works, 1899–1924, IN: ed. BOYDSTON, J.A., **Human nature and conduct**, Vol.14, Carbondale: Southern Illinois University Press, 1967.

DONAGAN, A., **The theory of morality**, Chicago: University of Chicago Press, 1977.

DWORKIN, R., **Taking rights seriously**, Cambridge: Harvard University Press, 1978.

ENGSTROM, S., **The form of practical knowledge: A study of the categorical imperative**, Cambridge, Mass.: Harvard University Press, 2009.

ENOCH, D. In defense of moral deference, **Journal of philosophy**, 111: 229–58, 2014.

FERNANDEZ, P. A., Practical reasoning: Where the action is, **Ethics**, 126: 869–900, 2016.

FLETCHER, J., **Situation ethics: The new morality**, Louisville: Westminster John Knox Press, 1997.

FRANKFURT, H. G., **The importance of what we care about: Philosophical essays**, Cambridge: Cambridge University Press, 1988.

GERT, B., **Morality: Its nature and justification**, New York: Oxford University Press, 1998

GIBBARD, Allan, Rule-utilitarianism: Merely an illusory alternative? **Australasian Journal of Philosophy**, 43: 211–220,1965.

GOLDMAN, Holly S., David Lyons on utilitarian generalization, **Philosophical Studies**, 26: 77–95, 1974.

GREENE, J. D., Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics, **Ethics**, 124: 695–726, 2014.

HABERMAS, J., **The theory of communicative action: Vol. I, Reason and the rationalization of society**, Boston: Beacon Press, 1984.

HAIDT, J., The emotional dog and its rational tail: A social intuitionist approach to moral judgment, **Psychological Review**, 108: 814–34, 2001.

HARE, R. M., **Moral thinking: Its levels, method, and point**, Oxford: Oxford University Press, 1981.
HARMAN, G., **Change in view: principles of reasoning**, Cambridge, Mass.: MIT Press, 1986.
HELD, V., **Justice, and care: Essential readings in feminist ethics**, Boulder, CO.: Westview Press, 1995.
HIERONYMI, P., The use of reasons in thought (and the use of earmarks in arguments), **Ethics**, 124: 124–27, 2013.
HORTY, J. F., **Reasons as defaults**, Oxford: Oxford University Press, 2012.
HORTY, J. F., Reasoning with precedents as constrained natural reasoning, in LORD E. and MCGUIRE, B. (ed.), **Weighing Reasons**, Oxford: Oxford University Press: 193–212, 2016.
HUME, D., [1739–40]. **A treatise of human nature**, ed. D. F. Norton and M. J. Norton, Oxford: Oxford University Press, 2000.
HURLEY, S. L., **Natural reasons: Personality and polity**, New York: Oxford University Press, 1989.
JONSEN, A. R., and TOULMIN S., **The abuse of casuistry: A history of moral reasoning**, Berkeley: University of California Press, 1988.
KAGAN, S., The additive fallacy, **Ethics**, 90: 5–31, 1988.
KNOBE, J., The concept of individual action: A case study in the uses of folk psychology, **Philosophical Studies**, 130: 203–231, 2006.
KOENIGS, M., Damage to the prefrontal cortex increases utilitarian moral judgments, **Nature**, 446: 908–911, 2007.
KOLODNY, N., Why be rational? **Mind**, 114: 509–63, 2005.
LADEN, A. S**., Reasoning: A social picture**, Oxford: Oxford University Press, 2012.
KORSGAARD, C. M., **Creating the kingdom of ends**, Cambridge: Cambridge University Press, 1996.
LANCE, M. and LITTLE, M., Where the Laws Are, in SHAFER-LANDAU, R. (ed.), **Oxford Studies in Metaethics** (Volume 2), Oxford: Oxford University Press, 2007.
LIST, C. and Pettit, P., **Group agency: The possibility, design, and status of corporate agents**, Oxford: Oxford University Press, 2011.
MCDOWELL, John, **Mind, value, and reality**, Cambridge, Mass.: Harvard University Press, 1998.
MCGRATH, S., The puzzle of moral deference**, Philosophical Perspectives**, 23: 321–44, 2009.

MCKEEVER, S. and RIDGE, M., **Principled Ethics: Generalism as a Regulative Idea**, Oxford: Oxford University Press, 2006.
MCNAUGHTON, D., **Moral vision: An introduction to ethics**, Oxford: Blackwell, 1988.
MILL, J. S., **Utilitarianism**, Indianapolis: Hackett Publishing, 1979 [1861].
MILLGRAM, E., **Practical induction**, Cambridge, Mass.: Harvard University Press, 1997.
MIKHAIL, J., **Elements of moral cognition: Rawls's linguistic analogy and the cognitive science of moral and legal judgment**, Cambridge: Cambridge University Press, 2011.
NELL, O., **Acting on principle: An essay on Kantian ethics**, New York: Columbia University Press, 1975.
NUSSBAUM, M. C., **Love's knowledge: Essays on philosophy and literature**, New York: Oxford University Press, 1990.
NUSSBAUM, M. C., **Upheavals of thought: The intelligence of emotions**, Cambridge, England: Cambridge University Press, 2001.
PIETROSKI, P. J., Prima facie obligations, ceteris paribus laws in moral theory, **Ethics**, 103: 489–515, 1993.
PRINZ, J., **The emotional construction of morals**, Oxford: Oxford University Press, 2007.
RACHELS, J., Active and passive euthanasia, **New England Journal of Medicine**, 292: 78–80, 1975
RAILTON, P., 1984. Alienation, consequentialism, and the demands of morality, **Philosophy and Public Affairs**, 13: 134–71.
RAILTON, P., The affective dog and its rational tale: Intuition and attunement, **Ethics**, 124: 813–59, 2014.
RAWLS, J., **A theory of justice**, Cambridge, Mass.: Harvard University Press, 1971
RAWLS, J., **Political liberalism**, New York: Columbia University Press, 1996.
RAWLS, J., **A theory of justice**, revised edition, Cambridge, Mass.: Harvard University Press, 1999.
RAWLS, J., **Lectures on the history of moral philosophy**, Cambridge, Mass.: Harvard University Press, 2000.
RAZ, J., **Practical reason and norms**, Princeton: Princeton University Press, 1990.
RICHARDSON, H. S., **Practical reasoning about final ends**, Cambridge: Cambridge University Press, 1994.
RICHARDSON, H. S., Specifying, balancing, and interpreting bioethical principles, **Journal of Medicine and Philosophy**, 25: 285–307, 2000.
RICHARDSON, H. S., **Democratic autonomy: Public reasoning about the ends of policy**, New York: Oxford University Press, 2002.

RICHARDSON, H. S., Thinking about conflicts of desires, in BAUMANN, P. and BETZLER, M. (ed.), **Practical conflicts: New philosophical essays**, Cambridge: Cambridge University Press, 96–117, 2004.

BETZLER, M. (ed.), **Articulating the moral community: Toward a constructive ethical pragmatism**, New York: Oxford University Press, 2018.

ROSS, W. D., **The right and the good**, Indianapolis: Hackett, 1988.

SANDEL, M., **Liberalism and the limits of justice**, Cambridge: Cambridge University Press, 1998.

SARTRE, J. P., Existentialism is a Humanism, in KAUFMANN, W. **Existentialism from Dostoyevsky to Sartre**, New York: Meridian-New American, 345–69, 1975.

SCHEFFLER, S., **Human morality**, New York: Oxford University Press, 1992.

SCHMIDTZ, D., **Rational choice and moral agency**, Princeton: Princeton University Press, 1995.

SCHNEEWIND, J.B., **Sidgwick's ethics and Victorian moral philosophy**, Oxford: Oxford University Press, 1977.

SCHROEDER, M., Holism, weight, and undercutting. **Noûs**, 45: 328–44, 2011.

SCHWITZGEBEL, E. and CUSHMAN, F., Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers, **Mind and Language**, 27: 135–53, 2012.

SIDGWICK, H., **The methods of ethics**, reprinted, 7th edition, Indianapolis: Hackett, 1981

SINNOTT-ARMSTRONG, W., **Moral dilemmas**, Oxford: Basil Blackwell, 1988.

SMITH, M., **The moral problem**, Oxford: Blackwell, 1994.

SMITH, M., A constitutivist theory of reasons: Its promise and parts, **Law, Ethics and Philosophy**, 1: 9–30, 2013.

SNEDDON, A. A social model of moral dumbfounding: Implications for studying moral reasoning and moral judgment**, Philosophical Psychology**, 20: 731–48, 2007.

SUGDEN, R., Thinking as a team: Towards an explanation of nonselfish behavior, **Social Philosophy and Policy**, 10: 69–89, 1993.

SUNSTEIN, C. R., **Legal reasoning and political conflict**, New York: Oxford University Press, 1996.

TIBERIUS, V., Humean heroism: Value commitments and the source of normativity, **Pacific Philosophical Quarterly**, 81: 426–446, 2000

VOGLER, C., Sex and talk**, Critical Inquiry**, 24: 328–65, 1998.

WELLMAN, H. and Miller, J., Including deontic reasoning as fundamental to theory of mind, **Human Development**, 51: 105–35, 2008.

WILLIAMS, B., **Moral luck: Philosophical papers 1973–1980**, Cambridge: Cambridge University Press, 1981.

WOLFF, R. P., **In defense of anarchism**, Berkeley and Los Angeles: University of California Press, 1998.

YOUNG, L.; SAXE, R., The neural basis of belief encoding and integration in moral judgment. **NeuroImage**, 40, p. 1912–1920, 2008.

(III) Motivação Moral*

Autor: Connie Rosati
Tradutor: Rafael Martins

* ROSATI, C. Moral Motivation. In: ZALTA, E. N. (ed.). The Stanford Encyclopedia of Philosophy. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2016. Disponível em: https://plato.stanford.edu/archives/win2016/entries/moral-motivation/. Acesso em: 15 jul. 2022.

The following is the translation of the entry on Moral Motivation by Connie Rosati, 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/win2016/entries/moral-motivation/. This translated version may differ from the current version of the entry, which may have been updated since the time of this translation. 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.

N.T.: A tradução da palavra overriding e suas variantes, overrdingly, override, overridingness, requer um esclarecimento, uma vez que o conceito de overridingness é muito utilizado em metaética. Ele se refere à propriedade que motivações, razões, ou juízos normativos, entre eles, os morais, possuem de superar, preponderar, ou prevalecer sobre motivações, razões, ou juízos de outras natureza, por exemplo, política, estética, prudencial, econômica, etc., em função de seu peso, grau, ou força normativa. As traduções para o verbo to override neste texto são: dominar, prevalecer, suplantar, superar, predominar. Para sua adjetivação ou nomeação, sugiro "suprema". (vide verbete Brink Supremacy, da SEF).
N.T.: Ao longo do verbete, as seguintes palavras ou expressões são traduzidas das seguinte forma: practicality requirement (condicional motivacional), practically rational (agencialmente racional, ou seja, que toma decisões e age baseado em razões), inverted commas (discurso indireto) e skin-conductive response (SCR) (resposta galvânica da pele ou resposta da atividade eletrodérmica).
Agradecemos a Roseline Crippa pela atenciosa revisão.

Em nossa vida cotidiana, enfrentamos uma série de questões morais. Uma vez que tenhamos deliberado e formado juízos sobre o que é **certo** ou **errado**, **bom** ou **mau**, esses juízos tendem a ter uma influência marcante sobre nós. Embora, no final, nem sempre nos comportemos como pensamos que devemos, nossos juízos morais normalmente nos motivam a agir de acordo com eles, pelo menos em algum grau. Quando os filósofos falam sobre motivação moral, este é o fenômeno básico que eles procuram compreender. A motivação moral é uma instância de um fenômeno mais geral, que podemos chamar de **motivação normativa** - pois nossos outros juízos normativos também costumam ter alguma força motivadora. Quando fazemos o juízo normativo de que algo é **bom para nós**, ou que temos uma **razão** para agir de uma maneira particular, ou que um curso de ação específico é **racional**, também tendemos a estar motivados. Muitos filósofos consideram a força motivadora dos juízos normativos como a característica principal que os marca como **normativos**, distinguindo-os, assim, de muitos outros juízos que fazemos. Em contraste com nossos juízos normativos, nossos juízos matemáticos e empíricos, por exemplo, parecem não ter nenhuma conexão intrínseca com motivação e ação. A crença de que um antibiótico vai curar uma infecção específica pode motivar um indivíduo a tomar o antibiótico, se ele também acreditar que tem a infecção e se ele desejar ser curado ou se julgar que deve tratar a infecção para seu próprio bem. Por si só, no entanto, uma crença empírica como essa parece não trazer consigo nenhum impacto motivacional particular; uma pessoa pode julgar que um antibiótico é mais eficaz para curar uma infecção específica sem ser movido de uma forma ou de outra.

Embora a força motivadora possa ser uma característica distintiva dos juízos normativos, o fenômeno da motivação normativa parece mais significativo no caso de juízos estritamente **morais**. Seja como for, a motivação moral recebeu muito mais atenção do que a motivação em relação a outros juízos normativos. Acredita-se amplamente que a moralidade entra em conflito, com frequência e às vezes severamente, com o que um agente mais valoriza ou prefere fazer. Talvez por causa da aparente oposição entre interesse próprio e moralidade, o fato da motivação moral parece especialmente intrigante. Como é que somos tão estavelmente movidos por nossos juízos morais? E qual é a natureza precisa da conexão entre juízo moral e motivação? É claro que quanto menos intrigante e mais comum a motivação moral nos parece, mais intrigantes se tornam as **falhas** de motivação moral. Se nosso objetivo é explicar a motivação moral, precisaremos compreender

não apenas como os juízos morais tão regularmente conseguem motivar, mas também como eles podem falhar em motivar, às vezes de maneira espetacular. Não apenas testemunhamos o fracasso motivacional entre pessoas com transtornos mentais, ou simplesmente abatidas e confusas, mas também entre pessoas saudáveis e em relativo controle de suas faculdades cognitivas e emocionais. O que devemos fazer com o **amoralista** - um indivíduo racional e motivado que aparentemente emite juízos morais, embora permaneça totalmente indiferente?

Ao responder às perguntas anteriores, os filósofos têm sido levados a visões nitidamente divergentes sobre a motivação moral, com implicações importantes para questões fundamentais em ética. Mais precisamente, as visões divergentes sobre a motivação moral envolvem o compromisso com teses particulares relacionadas a questões sobre a semântica moral e a natureza da moralidade. Talvez o mais popular seja o fato de certas teses terem sido desenvolvidas em conjunto para apoiar visões céticas ou **antirrealistas** em metaética. Este verbete fornece uma visão geral das principais posições que os filósofos assumiram em seus esforços para compreender e explicar o fenômeno da motivação moral. Neste verbete também se explica resumidamente como teses fundamentais relativas à motivação moral passaram a informar e estruturar debates sobre semântica moral e a natureza da moralidade.

1. O Fenômeno Básico da Motivação Moral

O fenômeno básico da motivação moral pode receber uma descrição mais sistemática da seguinte forma, usando 'P' para representar alguma pessoa e 'φ' e 'ψ' cada um para representar alguma ação:

> Quando P julga que φ seria a ação moralmente correta, ela normalmente está motivada para executar φ; se P mais tarde se convencer de que a ação φ seria errada, e que ao invés disso a ação ψ seja moralmente correta, esta pessoa normalmente deixa de estar motivada para φ e passa a ser motivada para ψ.

Esta representação visa capturar características de nossa experiência comum. Como a observação sugere, as pessoas geralmente se sentem motivadas a fazer o que julgam correto; além disso, sua motivação normalmente altera-se para

coincidir ou "rastrear" mudanças em seus juízos morais. Se uma pessoa julgar certo cumprir uma promessa em vez de ajudar um estranho em necessidade, ela normalmente se sentirá motivada, pelo menos em algum grau, a agir de modo a cumprir a promessa. Se ela mudar de ideia sobre a prioridade de sua promessa, normalmente deixará de estar motivada a cumprir a promessa e, em vez disso, passará a estar motivada a fornecer ajuda.

Antes de abordarmos as muitas questões que a apresentação acima deixa em aberto, e que estão no cerne dos debates sobre a natureza da motivação moral, devemos observar dois pontos importantes. Primeiro, a apresentação acima não diz nada sobre a **força** da motivação moral. Por tudo o que nos diz, a motivação que todas ou algumas pessoas sentem para fazer o que julgam certo pode ser extraordinariamente fraca. A experiência comum sugere que a motivação moral de fato tende a ser bastante robusta, mas com uma ressalva a ser observada mais tarde, as visões filosóficas sobre a motivação moral geralmente seguem o costume de não se tomar posição quanto à força exata da motivação moral. Em segundo lugar, a apresentação reflete uma suposição amplamente compartilhada, que faz parte do pano de fundo para debates sobre a natureza da motivação moral, a saber, que a motivação moral é um **fenômeno surpreendentemente regular e confiável**. Ao longo da vida social, tanto em nossas relações pessoais quanto em nossas interações públicas, consideramos como certo que os juízos morais motivam de forma confiável, senão infalível, e que efetivamente influenciam e orientam como as pessoas se sentem e agem. Ainda assim, a suposição não é totalmente incontroversa; na verdade, alguns expressam sérias dúvidas sobre se a motivação moral é tão regular e confiável como comumente supomos (COPP, 1997, p. 50).

O fenômeno básico da motivação moral parece relativamente simples. A difícil tarefa filosófica é tentar compreender e explicar mais completa e precisamente a natureza da motivação moral. As Seções 2 e 3 exploram duas abordagens para a tarefa. Embora a abordagem discutida na seção 3 tenha sido predominante, a abordagem a ser considerada brevemente na seção 2 fornece um contraste instrutivo, bem como um primeiro vislumbre de como as teses sobre motivação moral se relacionam com questões metaéticas mais amplas. A seção 4 explora considerações mais gerais sobre motivação moral e metaética, enquanto a seção 5 considera supostas implicações para teorias filosóficas sobre motivação moral a partir de trabalhos recentes em psicologia empírica.

2. Motivação Moral e a Natureza das Propriedades Morais

Quando julgamos que uma ação é certa ou errada ou que um estado de coisas é bom ou ruim, parecemos representar o mundo como sendo de uma determinada maneira. Parece que expressamos uma crença moral, atribuindo uma propriedade moral particular ou característica normativa às ações ou estado de coisas. Tomando como guia a aparente forma representacional dos juízos morais, podemos tentar explicar a motivação moral apelando para a natureza das **propriedades** que figuram em nossos juízos morais. Talvez sejamos motivados de forma confiável por nossos juízos morais, pelo menos quando esses juízos são aproximadamente corretos, porque as próprias propriedades morais como **correção** e **bondade** nos motivam, quando as apreendemos.

J. L. Mackie (1977) faz uma crítica famosa a essa imagem das propriedades morais em seu amplo argumento contra a objetividade da ética. Mackie afirma ter encontrado algo parecido com tal objetividade no trabalho de várias figuras históricas, incluindo Kant e Sidgwick, mas sua apresentação mais clara da imagem vem em suas observações sobre Platão. Mackie escreve:

> Na teoria de Platão, as Formas, e em particular a Forma do Bem, são realidades eternas, extramentais. Elas são um elemento alicerçador central na estrutura do mundo. E mesmo assim, sustenta-se que apenas conhecer ou "ver" estas formas não apenas dirá aos homens o que fazer, mas também garantirá que o façam, prevalecendo sobre quaisquer inclinações contrárias. Os reis-filósofos da *República* podem, pensa Platão, ser confiados com poder irrestrito porque sua educação os terá dado conhecimento das Formas. Estando familiarizados com as Formas do Bem, da Justiça e da Beleza e outras, eles serão, somente por esse conhecimento, sem qualquer motivação adicional, impelidos a perseguir e promover esses ideais. (MACKIE, 1977, p. 23-24)

Certas características da tese de Platão sobre a motivação moral - ou pelo menos a caracterização de Mackie dela - merecem atenção. Em primeiro lugar, como Mackie interpreta a visão de Platão, a motivação moral surge direta e

inteiramente da compreensão da presença das próprias propriedades morais. A apreensão dessas propriedades move um agente a agir, sem a ajuda de qualquer fonte adicional de motivação; seu poder motivacional não depende de nenhum desejo ou disposição do próprio indivíduo. Em segundo lugar, a apreensão de propriedades morais não apenas motiva por si mesma: ela fornece uma motivação **preponderante**. Uma vez que um agente apreende tais propriedades, seu poder motivador supera quaisquer desejos ou inclinações opostas.

Quando Mackie mantem que a teoria das Formas de Platão descreve como seriam os valores objetivos, ele na verdade subscreve (e atribui a Platão) uma visão chamada **internismo existencial**. De acordo com o internismo existencial, existe uma conexão necessária entre algo ter um certo status normativo e motivação.[40] Por exemplo, um estado de coisas não poderia *ser* bom, a menos que sua apreensão fosse capaz de motivar, embora não precise motivar de forma preponderante. Se um indivíduo apreende algo e não age de acordo, então, *ceteris paribus*, este algo não é bom. Porém, conforme Mackie descreve a visão de Platão, os valores objetivos fornecem uma motivação preponderante e, portanto, a visão reflete uma versão particularmente forte de internismo existencial. De acordo com essa versão, um estado de coisas não seria bom ou valioso, a menos que sua apreensão fornecesse uma motivação preponderante. O caráter internista da imagem platônica de Mackie curiosamente a alinha com as visões contemporâneas que da mesma forma aceitam formas de internismo existencial, enquanto sustentam que a capacidade de motivação na verdade depende de um desejo preexistente. Considere uma visão sobre razões para agir associadas mais proeminentemente a Bernard Williams (1981). De acordo com o que é chamado de **internismo de razões** ou **internismo racional**, necessariamente, se um indivíduo tem uma razão para agir, ele deve ser capaz de estar motivado para assim agir. Porém, na visão de Williams, para estar motivado, um indivíduo precisa ter alguma atitude motivadora em seu "conjunto motivacional" atual. Portanto, grosso modo, se uma consideração não motiva uma pessoa de acordo com seus desejos atuais ou conjunto motivacional atual, esta consideração não pode ser uma razão para agir. Tanto a perspectiva de Williams, quanto a interpretação que Mackie faz de Platão, postulam uma conexão necessária entre status normativo e motivação, embora Williams faça o status normativo depender de elementos subjetivos de um indivíduo, o que a visão Platônica rejeitaria categoricamente.

---

[40] *Vide* Darwall (1983, p. 51-). Confira, também, as notas 8 e 10.

A discussão de Mackie fornece uma primeira ilustração de como teorias da motivação moral tem sido empregadas para defender ou refutar posições mais amplas em metaética. De acordo com Mackie, o poder motivador dos valores objetivos, se estes existissem, teria de ser exatamente como Platão o descreveu.

> As Formas, de Platão, dão uma imagem dramática de como os valores objetivos teriam de ser . A Forma do Bem é tal que o conhecimento dela fornece ao conhecedor uma direção e um motivo preponderante a bondade de algo instrui a pessoa que a apreende a persegui-la e a faz perseguir. Um bem objetivo seria buscado por qualquer um que o conhecesse, não por causa de qualquer fato contingente acerca de como uma pessoa, ou qualquer pessoa, é constituída de tal forma que deseja esse fim, mas apenas porque o fim tem, de alguma forma embutida nele mesmo, a propriedade de incitar sua perseguição. (MACKIE, 1977, p. 40)

Mackie defende que as sentenças morais que proferimos quando fazemos juízos morais na verdade expressam proposições sobre exatamente essas propriedades "objetivamente prescritivas"; como resultado, nossos juízos morais podem ser verdadeiros ou falsos. Assim, o **cognitivismo moral** - a visão de que juízos e crenças morais, e as sentenças que os expressam, podem ser verdadeiras ou falsas - fornece a explicação correta da semântica moral, do que significam nossos juízos morais. Dado que nosso discurso moral é **cognitivista**, somos induzidos a presumir a autenticidade do **realismo moral**, a visão, grosso modo, de que juízos e crenças morais possuem valor de verdade, e alguns deles são literalmente verdadeiros.[41] Mas o discurso moral sofre do que Mackie chama de "falha de pressuposição": o discurso moral pressupõe propriedades objetivamente prescritivas, embora elas não existam; tais propriedades teriam que ser "entidades estranhas",

---

[41] Esta forma de explicar o realismo moral segue Sayre-McCord, em *The Many Moral Realisms*. (*In*: SAYRE-MCCORD, G. **Essays on Moral Realism**. Ithaca: Cornell University Press, p. 1-23).

diferentes de qualquer outra coisa no mundo. Evidentemente, Mackie pensa que falar sobre moralidade é mais como falar sobre unicórnios. Nossa "conversa de unicórnio" expressa (pelo menos assumindo a lenda medieval) proposições sobre criaturas semelhantes a cavalos, domesticáveis apenas por virgens, cujos chifres espirais possuem poderes mágicos. Mas tais criaturas não existem e, portanto, nosso discurso sobre unicórnios está sistematicamente errado, mesmo que alguns de nós já não incorramos neste erro. Ao negar a existência de propriedades morais, Mackie rejeita o realismo moral, combinando uma semântica moral cognitivista com uma **teoria do erro**. De acordo com a teoria do erro, "embora a maioria das pessoas, ao fazer juízos morais, afirme implicitamente [...] estar apontando para algo objetivamente prescritivo, essas afirmações são todas falsas" (MACKIE, 1977, p. 35).[42]

Embora os filósofos contemporâneos divirjam com relação ao ceticismo moral de Mackie, eles concordam principalmente em rejeitar suas afirmações extremamente fortes sobre como devem ser a motivação moral e as propriedades morais objetivas que figuram em nossos juízos morais. Eles uniformemente rejeitam a sugestão de que uma compreensão das obrigações da moralidade produziria uma motivação preponderante para agir em conformidade com suas demandas. A maioria rejeita os esforços para explicar a motivação moral apelando para um poder motivador que emana das propriedades morais e dos atos e estados de coisas que as instanciam. Uma exceção limitada a esta última afirmação é digna de nota. Christine Korsgaard (1996) endossa a ideia de algo como entidades objetivamente prescritivas, embora essas entidades não sejam, em sua opinião, propriedades morais. Korsgaard compartilha do ceticismo de Mackie sobre os valores objetivos do tipo que ele descreve como figurando nas teorias morais realistas de filósofos como Platão. No entanto, ela observa, Mackie está enganado e o realista está certo com relação a se alguma entidade existente pode atender ao duplo critério de fornecer ao agente que a conhece "uma direção e um motivo". Ela insiste: "é o fato mais familiar da vida humana que o mundo contém entidades que podem nos dizer

---

[42] As afirmações de Mackie, sobre as propriedades morais, não são inteiramente claras, mas sua posição parece ser, não apenas que nada possui propriedades morais, mas que propriedades morais não existem. Na medida em que Mackie está comprometido com a inexistência de propriedades morais, a analogia com os unicórnios é pelo menos um tanto inadequada. Pois é possível argumentar que a propriedade de ser um unicórnio exista, embora não seja instanciada, em outras palavras nenhum objeto a possui.

o que fazer e nos obrigar a fazê-lo. Eles são pessoas e os outros animais" (KORSGAARD, 1996, p. 166). Entretanto, a maioria dos filósofos, mesmo aqueles que simpatizam com a filosofia moral de Kant e com o tipo de kantismo de Korsgaard, consideram ilusória a ideia de que as pessoas (e os animais não humanos) possuem valor e podem, a esse respeito, "nos dizer o que fazer" e "nos obrigar a fazê-lo". Mas as afirmações de Korsgaard são parte de uma extensa e extremamente rica caracterização da ética que não pode ser explorada aqui, pois uma avaliação justa de suas afirmações exigiria atenção a esse quadro mais amplo. O ponto importante, para os presentes propósitos, é que pelo menos alguns filósofos, Korsgaard, e talvez outros atraídos por ideias derivadas da filosofia moral de Kant, são atraídos para a ideia de que a motivação moral e a normatividade encontram sua fonte em entidades inerentemente normativas ou "objetivamente prescritivas".

Quer existam ou não quaisquer propriedades ou entidades com alguma forma dos poderes descritos por Mackie, é um erro supor que os realistas e os objetivistas morais precisam estar comprometidos com sua existência. Nenhum realista ou objetivista precisa pensar que as propriedades morais, ou fatos sobre sua instanciação, serão, quando apreendidos, suficientes para motivar todas as pessoas independentemente de suas circunstâncias, incluindo sua constituição cognitiva e motivacional. E os realistas certamente não precisam assumir a visão que Mackie atribui a Platão, de que a apreensão de valores objetivos garantirá que alguém aja, "prevalecendo sobre qualquer inclinação contrária" (MACKIE, 1977, p. 23). Um indivíduo pode compreender um fato moral, mas pode, por exemplo, sofrer de irracionalidade temporária ou fraqueza de vontade; ele pode ainda estar livre de tais defeitos temporários, mas possuir uma composição motivacional mais refratária que impede ou derrota o poder motivador dos fatos morais. Qualquer explicação plausível da motivação moral irá, e deve, reconhecer essas fontes de falha motivacional; e qualquer análise plausível das propriedades morais deve também as incluir. Nem mesmo aqueles realistas ou objetivistas que afirmam que todas as pessoas **racionais** e **motivacionalmente íntegras** serão movidas por fatos morais, precisam pensar que elas serão motivadas de forma prevalecente e irrevogável. Como já observado, independentemente de seus pontos de vista com respeito a questões metaéticas mais amplas, os filósofos contemporâneos não assumem qualquer posição sobre a força precisa da motivação moral - com a ressalva (aludida anteriormente) de que rejeitam, aparentemente de forma universal a ideia de que a motivação moral é comumente preponderante.

3. Juízo Moral e Motivação

Na maioria das vezes, os filósofos tentaram explicar a motivação moral não apelando para os poderes especiais das propriedades morais, mas para a natureza dos **juízos morais**. Talvez os juízos morais possuam uma natureza tal que nenhuma pessoa poderia julgar sinceramente um ato moralmente correto ou uma situação moralmente boa, e permanecer **totalmente** impassível. Os esforços para compreender a motivação moral em termos de motivação por juízos morais devem confrontar duas questões centrais. Em primeiro lugar, qual é a natureza da conexão entre juízo moral e motivação - os juízos morais motivam **necessariamente** ou motivam apenas **contingentemente**? Em segundo lugar, os juízos morais podem motivar por si próprios ou eles podem motivar apenas pela intermediação de um desejo ou outro estado conativo? Evidentemente, os filósofos têm respondido estas questões de maneiras variadas.

3.1. Humianismo *vs*. Anti-Humianismo

Primeiro vamos considerar a segunda pergunta. Uma maneira pela qual os juízos morais poderiam motivar, e, de fato, motivar por conta própria, seria se afinal os juízos morais não fossem representacionais. Suponha que os juízos morais não atribuam propriedades nem expressem crenças morais sobre quais coisas possuem essas propriedades. Suponha que, em vez disso, como afirma o **não-cognitivismo moral**, que os juízos morais expressem desejos ou outros estados **conativos** - o que os filósofos às vezes chamam de "pró-atitudes". Desta forma ficaria claro como os juízos morais se conectam à motivação. Eles simplesmente expressam um estado motivador que o indivíduo já possui; fazer um juízo moral (sincero) já é estar motivado, pelo menos até certo ponto. O verdadeiro enigma sobre como os juízos morais podem motivar surge para aqueles que sustentam que estes juízos expressam **crenças morais**, pois a conexão entre a crença, um estado **cognitivo**, e a motivação é incerta.

O modo como os filósofos resolvem o desafio gira em torno de uma questão central na psicologia moral, a saber, se o que é chamado de **teoria humiana** da motivação é verdadeira. Na visão humiana, a crença é insuficiente para a motivação, que requer sempre, além da crença, a presença de um desejo ou estado conativo. A motivação moral, portanto, não pode surgir apenas da crença moral, mas deve

depender também de um desejo preexistente ou outro estado conativo ou intrinsecamente motivador. Talvez seja justo dizer que o humianismo continua a ser a visão dominante. Ele tem sido sustentado não apenas por aqueles que rejeitam o cognitivismo e o realismo moral, mas também por alguns que aceitam estas teorias, de modo que, sozinho, o humianismo não é considerado decisivo para resoluções de questões mais amplas em metaética. Esta tradição tem sido defendida por não-cognitivistas antirrealistas, mas, também, por realistas morais, como Michael Smith (1994) e Peter Railton (1986a). Entretanto, vários filósofos proeminentes incluindo Thomas Nagel, (1970), John McDowell (1979), Mark Platts (1980), David McNaughton (1988), Jonathan Dancy (1993), Thomas Scanlon (1998), and Russ Shafer-Landau (2003), têm rejeitado a perspectiva humiana argumentando que a motivação moral na verdade não depende da existência do desejo: crenças morais podem, por si só, dar origem a motivação.

Precisamente como e em quais condições uma crença moral pode motivar é uma questão controversa entre os anti-humianos. Alguns sustentam que a crença moral é suficiente para motivar diretamente. O mero fato de acreditar que algo é certo, digamos, cumprir uma promessa, moverá a pessoa, pelo menos até certo ponto, a agir de modo a cumprir a promessa. Outros sustentam que as crenças morais produzem desejos, que por sua vez nos motivam em conjunto com as crenças morais que os produziram. Acreditar que é certo cumprir uma promessa produz o desejo de cumpri-la, e esses estados cognitivos e conativos conjuntamente movem o indivíduo, pelo menos em algum grau, a agir de modo a manter a promessa. Certos teóricos da ética das virtudes oferecem uma versão bastante refinada da última ideia, argumentando que apenas um tipo particular de crença moral - aquela ligada a um ideal ou concepção completa de uma situação à luz de uma compreensão mais ampla sobre como viver - necessariamente gera em um indivíduo a motivação para agir da forma como a crença moral relevante indica que ela deveria (LITTLE, 1997; MCDOWELL, 1978). A pessoa virtuosa não tem meras crenças morais, mas um complexo de crenças e perspectivas morais que com segurança a moverão a agir moralmente. Os proponentes de vários pontos de vista anti-humianos reconhecem prontamente que as pessoas muitas vezes deixam de agir como acreditam que deveriam. De acordo com qualquer uma dessas visões, entretanto, a falha de motivação surge de uma falha cognitiva.

Como já foi observado, muitos consideram a estrutura humiana básica mais plausível. Antes de examinar algumas das considerações favoráveis a ela, devemos notar o fato de que o próprio humianismo não compromete ninguém com

nenhuma visão particular quanto aos tipos de desejos responsáveis pela motivação moral. Um humiano pode muito bem considerar que nenhum desejo particular está implicado na motivação moral. Pelo contrário, desejos variados podem, quando contingentemente presentes, induzir um indivíduo a fazer o que ele julga que deve fazer, incluindo o desejo de ser bem-visto por seus vizinhos, de defender seus interesses de alguma forma, ou de promover o bem-estar daqueles que importam para ela. Entretanto, apelar simplesmente para um ou outro desejo contingente pode ser inadequado para explicar o fenômeno básico da motivação moral. Afinal, o que precisa ser explicado, muitos argumentam, não é apenas como podemos, por ocasião ou mesmo com frequência, ser motivados a fazer o que pensamos que devemos: o que precisa ser explicado é como somos **regularmente** motivados a fazer o que pensamos que devemos. Isso inclui explicar por que a motivação se **reorienta** de maneira regular, de modo a acompanhar mudanças em nossas crenças morais. Como veremos, aqueles que aceitam o quadro humiano às vezes sugerem que olhemos para desejos bastante particulares ou para características profundas da psicologia humana para explicar a motivação moral.

Um argumento a favor do quadro humiano alega que, se as crenças fossem suficientes para motivar, seria de se esperar que as pessoas com as mesmas crenças fossem motivadas da mesma maneira. Entretanto, na verdade, enquanto algumas pessoas são motivadas por suas crenças morais, por exemplo, que contribuir para o alívio da fome é um dever, e, portanto, põe-se a preencher um cheque para a Oxfam, outras não sentem tal inclinação de forma alguma. Não obstante, os anti-humianos afirmam que podem explicar essas diferenças mostrando que a diferença motivacional na verdade se deve a outras diferenças de crença ou a motivos que competem com os desejos gerados por crenças morais e os suplantam. (SHAFER-LANDAU, 2003, p. 129-130)

Um segundo argumento a favor do humianismo apela à teoria das razões associada a Williams (1981), brevemente discutida anteriormente. Recordemos que de acordo com o internismo sobre razões ou internismo racional, se um indivíduo tem uma razão para agir, então ele necessariamente deve ser capaz de estar motivado para executar essa ação. Em uma versão mais específica da teoria, um indivíduo tem uma razão para agir apenas se ele possui o desejo de executá-la ou alcançar algum fim que exija a realização dessa ação. Se o internismo racional estiver correto, então, quando um indivíduo julga corretamente ter uma razão para agir, ele já deve ter um desejo preexistente. Os anti-humianos às vezes rejeitam o internismo de razões, bem como a teoria humiana da motivação. Mas mesmo

admitindo que o internismo racional esteja correto, eles acreditam que este segundo argumento falha em minar sua posição. Pois parece possível que nem todos os nossos juízos morais envolvam a conclusão (correta ou não) de que temos uma razão para agir. Por exemplo, uma pessoa poderia julgar que seria correto cumprir uma promessa sem concluir que ela tenha uma razão para assim agir. O que pode explicar isso? Talvez, por exemplo, ela deixe de refletir sobre a conexão entre o que é certo fazer e o que se tem razão para fazer; ou talvez ela acredite erroneamente que as verdades sobre ações moralmente corretas não implicam verdades sobre o que alguém tem razão para fazer. Se um indivíduo pode julgar uma ação correta sem julgar que ela tem uma razão para realizar a ação, então, mesmo que o fato de uma ação ser certa envolva uma razão para a ação e as razões envolvam desejos, as crenças morais não precisam envolver desejos preexistentes (SHAFER-LANDAU, 2003, p. 128-129)[43] .

Talvez o argumento mais sofisticado em favor da teoria humiana da motivação seja aquele que apela para considerações oriundas da filosofia da mente e psicologia moral, especificamente, para diferenças fundamentais entre crença e desejo, que contariam contra o anti-humianismo. Em âmbito conceitual, argumenta-se que crença e desejo diferem no que tem sido chamado de suas “direções de ajuste” (ANSCOMBE, 1963). Ao que parece, eles diferem de tal maneira que estados mentais de crença não podem implicar estados mentais de desejo. Enquanto as crenças têm como objetivo se adequar ao mundo, os desejos visam mudar o mundo. Ou seja, enquanto as crenças têm uma direção de ajuste “mente-mundo”, os desejos têm uma direção de adequação “mundo-mente”. Para que um estado mental conte como uma crença, ele deve ser em alguma medida reativo às evidências relativas à verdade ou falsidade de seu conteúdo proposicional; fatos contrários a uma crença contam contra ela. Em contraste, os fatos contrários ao conteúdo proposicional de um desejo - o fato de que o mundo não é realmente como se deseja - não precisam ser considerados evidência contra esse desejo. Precisamente porque os desejos não visam responder ao mundo, mas fazer com que o mundo responda a eles (para

---

[43] Para uma apresentação mais refinada deste e de outros argumentos humianos, bem como respostas anti-humianas mais detalhadas, confira Shafer-Landau (2003, p. 122-141). Shafer-Landau apresenta esses argumentos como considerações que um humiano pode oferecer para minar o que o autor considera ser uma "presunção a favor do anti-Humianismo." (p. 127).

fazer o mundo se ajustar a seus conteúdos proposicionais ou ao objeto do desejo), eles podem persistir mesmo quando o mundo se recusa a cooperar. O argumento [em favor da teoria humiana da motivação] diz que, assumindo que a teoria das direções de ajuste sobre crença e desejo seja verdadeira, pelo menos algumas versões de anti-humianismo exigiriam algo incoerente, a saber, estados mentais com direções de ajuste incompatíveis: estados mentais que poderiam ser ao mesmo tempo representacionais como as crenças e motivacionais como os desejos. Porém, os anti-humianos argumentam que sua teoria da motivação moral por meio da crença moral não precisa envolver tal incoerência. Para entender este ponto, precisamos apenas considerar a possibilidade de que um estado mental **poderia** ter direções de ajuste opostas, desde que, ao exibir cada direção de ajuste, o estado mental fosse direcionado a diferentes proposições: por exemplo, o agente virtuoso "acredita" (direção de ajuste de crença) *que um estado de coisas S deve ser promovido* e "deseja" (direção de ajuste de desejo) *que S seja realizado*. (LITTLE, 1997, p. 64)[44]

Os anti-humianos oferecem várias considerações - algumas positivas, outras negativas - para apoiar sua rejeição ao humianismo. No lado negativo, eles tentam derrubar considerações em favor da teoria humiana, como já vimos ao explorar algumas delas. No lado positivo, os anti-humianos às vezes invocam a fenomenologia da motivação moral, argumentando que ela sustenta sua visão. Pergunte a uma pessoa por que ela agiu em última instância conforme acreditava que a moralidade exigia enquanto terrivelmente tentada a fazer o contrário, e ela não relatará seus desejos no momento da ação; em vez disso, ela explicará que acreditava que a ação era “a coisa certa a fazer” (SHAFER LANDAU, 2003, p. 123). Nossa própria experiência e a de outras pessoas nos dizem que, embora nossas ações frequentemente surjam de nossos desejos, às vezes surgem de nossas crenças avaliativas. Como suporte adicional para essas afirmações sobre a fenomenologia da motivação moral, Shafer-Landau apela para casos não-morais em que a motivação parece decorrer da crença. Considere um indivíduo que se convence de que tem um desejo que de fato não tem, como o desejo de se tornar um advogado. Ele se matricula na faculdade de direito apenas para descobrir-se

---

[44] Little apresenta esta resposta no contexto da defesa dos teóricos da virtude. Veja Little (1997), de forma mais geral, para um exame extenso e detalhado dos argumentos contra a teoria da virtude com base no apelo às considerações da filosofia da mente.

desmotivado por seus estudos e abandona a escola. Depois de passar um verão como carpinteiro descobre seu amor pela carpintaria (SHAFER-LANDAU, 2003, p. 125). O que mais plausivelmente explica sua entrada na faculdade de direito e seus tímidos esforços durante aquele primeiro ano parece ser sua crença equivocada de que desejava se tornar um advogado. Visto que muitas de nossas escolhas envolverão nos sujeitarmos a experiências tediosas, e até dolorosas - experiências que certamente nenhum de nós deseja como um fim - o humiano nos deve alguma explicação sobre nossa a disposição de persistir em tais escolhas. Os humianos, ao que parece, são forçados a apelar para algum outro desejo que, com esta ação, buscamos satisfazer, como, por exemplo, o desejo de se tornar um advogado, no caso do abandono da faculdade de direito. Mas tal explicação será implausível nos casos em que estamos enganados sobre nossos desejos. Nenhuma justificativa convincente pode ser dada para aceitarmos uma explicação de nossas ações baseada no desejo, argumenta Shafer-Landau em defesa de uma explicação mais direta em termos de nossas crenças.

Ainda assim, os humianos insistem que não há nada direto nas tentativas de explicar a motivação moral e as ações em termos de crenças; basta relembrar o argumento a favor do humianismo baseado nas diferenças de "direção de ajuste" entre a crença e o desejo. Deixando esse argumento de lado, entretanto, nem a fenomenologia da motivação moral, nem os casos em que os indivíduos se enganam sobre seus desejos sustentam a visão anti-humiana. O fato de um indivíduo citar uma crença em vez de um desejo ao explicar por que fez o que julgou ser correto nada faz para mostrar que sua crença moral a moveu diretamente a agir ou que gerou um desejo que a levou a agir. Notoriamente, autorrelatos individuais não são confiáveis e, dificilmente, podem resolver uma questão tão fundamental sobre psicologia moral. Quanto aos casos em que os indivíduos estão (supostamente) equivocados sobre seus desejos, o senso comum sugere que o humiano tem uma explicação mais direta. O humiano pode argumentar que o desistente da faculdade de direito de fato desejava se tornar um advogado, ou pelo menos se matricular na faculdade de direito; ele simplesmente não entendia como seria estudar direito. Depois de vivenciar esta experiência, ele perdeu o desejo de continuar seus estudos. Alternativamente, talvez ele realmente não desejasse se tornar um advogado, embora ele dissesse a si mesmo que sim. Ainda assim, ele foi movido a entrar na faculdade de direito não meramente por sua crença, mas por um desejo mais profundo, talvez não totalmente consciente, como o desejo de agradar seus pais ou de ter o prestígio ou o retorno financeiro que acompanha a advocacia. Os anti-

humianos não nos deram nenhuma razão para favorecer sua explicação sobre as alternativas humianas. É claro que os anti-humianos não precisam pensar que a fenomenologia resolve a disputa, como eles supõem, mas os humianos insistirão que ela nem mesmo tende a favorecer a posição anti-humiana.

É claro que a discussão anterior não cobre todos os argumentos apresentados ao longo do debate entre humianos e anti-humianos, mas apenas alguns dos que os filósofos têm evidentemente considerado mais convincentes. Se, e como, o debate pode ser resolvido permanece incerto, em parte, porque a natureza da disputa é um tanto obscura. Seria, no fundo, uma disputa conceitual a ser resolvida, por exemplo, pela análise dos conceitos de crença e desejo? É possível, embora os argumentos que apelam a considerações da filosofia da mente e da psicologia moral tenham se mostrado menos do que totalmente convincentes. Em vez disso, seria a disputa fundamentalmente empírica? A tendência de apelar para o senso comum e a fenomenologia da ação moral parece trair qualquer tentação de tratar a questão como empírica, pelo menos em parte, embora talvez esses apelos tenham como objetivo servir apenas como um controle sobre as afirmações conceituais. Em qualquer caso, os apelos à nossa experiência podem ser igualmente, e de forma inconclusiva, invocados por aqueles em ambos os lados do debate. No contexto de afastar críticas à visão de que virtude é conhecimento, Little (1997) sugere que a disputa é fundamentalmente teórica, envolvendo grandes e complexas questões sobre a natureza da agência, normatividade e responsabilidade. Quer isto esteja correto ou não, Little pode estar certo ao sugerir que a disputa não será resolvida apelando para argumentos meramente locais do tipo que consideramos. O quão plausível alguém avalia ser qualquer um dos lados pode depender, no final, da plausibilidade das teorias mais amplas em que essas visões, respectivamente, figuram.

3.2 Internismo vs. Externismo

Seja o que for que se possa concluir sobre se os juízos ou crenças morais motivam por conta própria ou apenas por meio de algum estado conativo preexistente, uma questão permanece quanto à natureza precisa da conexão entre o juízo moral e a motivação. Os juízos morais motivam **necessariamente** ou motivam apenas **contingentemente**? Se for contingente, então como explicar por que a conexão contingente entre o juízo moral e a motivação é tão forte e confiável quanto parece?

O principal dissenso a respeito da natureza da conexão entre juízo moral e motivação se dá entre os filósofos que aceitam e aqueles que rejeitam uma tese conhecida como **internismo motivacional**. Esta tese é uma forma de **internismo adjudicatório**. Tradicionalmente, o internismo adjudicatório é caracterizado como estabelecendo que a motivação é interna ao juízo moral, no sentido de que o próprio juízo moral motiva sem a necessidade de um desejo acompanhante ("internismo forte") ou que há uma conexão necessária entre o juízo moral e a motivação ("internismo fraco"). Conforme caracterizado atualmente na literatura, o internismo adjudicatório afirma que, no âmbito conceitual, existe uma conexão necessária entre um juízo moral sincero e as razões ou motivos justificatórios: necessariamente, se um indivíduo julga sinceramente que deve φ, então ela tem uma razão ou motivo para φ. O internismo adjudicatório deve ser distinguido da tese do internismo existencial, que consideramos anteriormente. Lembremos que de acordo com o **internismo existencial**, existe uma conexão necessária entre algo ter um certo status normativo e motivação.[45] Por exemplo, uma consideração pode **ser** uma razão ou **atribuidora** de propriedades morais, por exemplo, apenas se for capaz de motivar. Enquanto o internismo adjudicatório declara uma condição necessária para que um juízo pertença a uma certa classe, o internismo existencial estabelece uma condição necessária para que um ato, um estado de coisas ou uma consideração pertença a uma certa classe normativa.

O internismo motivacional sustenta ser impossível que uma pessoa expresse um juízo moral sincero sem estar, pelo menos em algum grau, **motivada** a agir de acordo com este juízo. O internismo pode assumir formas mais fracas ou mais fortes. No relato de Mackie sobre Platão, temos a visão de que, necessariamente, uma pessoa que faz um juízo moral sincero terá uma motivação **prevalecente** para cumprir seu juízo. Logo, propriedades morais objetivas devem ser algo que envolve uma forma bastante extrema de internismo existencial, que seria aliada a uma forma

---

[45] Darwall (1983, p. 51), originalmente, cunhou as expressões "internismo adjudicatório" e "internismo existencial". Ele descreve o internismo adjudicatório sustentando que é uma condição necessária para que um juízo seja genuinamente de um certo tipo – que seja, digamos, o juízo que se deve executar φ ou o juízo que se tem uma razão para executar φ - que o indivíduo que faz o julgamento esteja disposto a agir de acordo com ele. Para uma distinção relacionada, embora ligeiramente diferente, entre os tipos de internismo, confira Brink (1989, p. 37-80).

bastante extrema de internismo adjudicatório. Os filósofos morais contemporâneos não são mais atraídos por uma afirmação tão forte quando a motivação moral é ligada ao julgamento moral do que o são quando a motivação moral é ligada a propriedades morais. Em vez disso, eles aceitam formas mais fracas de internismo, onde se permite que, embora, necessariamente, a pessoa que faz um juízo moral sincero sinta **alguma** motivação para cumpri-lo, essa motivação pode ser superada por desejos conflitantes e abatida por uma variedade de distúrbios mentais, como depressão e fraqueza de vontade. (SVAVARSDOTTIR, 1999, sec.1)

Como já deve estar evidente, aqueles que aceitam uma ou outra forma de internismo motivacional têm uma explicação disponível da estabilidade da motivação moral, incluindo a estabilidade com que a mudança motivacional acompanha mudanças no juízo moral. De fato, um argumento oferecido em favor do internismo é que somente se o aceitarmos podemos explicar de maneira plausível por que as mudanças na motivação moral seguem de forma confiável as mudanças no juízo moral (SMITH, 1994, p. 71-76). Suponha que Jones e Thomson estejam debatendo a permissibilidade moral do aborto. Jones tende a acreditar que o aborto é moralmente errado. Ela é conhecida por se juntar à fila de protesto em frente a uma clínica de aborto local e por tentar dissuadir as mulheres de fazerem abortos. Thomson, em contraste, acredita que o aborto é moralmente permissível. Suponha que, após extensa discussão, Thomson convença Jones de que os argumentos mais plausíveis sustentam a permissibilidade do aborto. O que se poderia razoavelmente prever em termos da conduta futura de Jones? Entre outras coisas, poderíamos razoavelmente prever que Jones não estaria mais inclinada a se juntar à fila de protesto e que desistiria de seus esforços para desencorajar outras mulheres a fazerem abortos. Mas essa previsão repousa precisamente na expectativa de que, pelo menos na medida em que Jones é uma pessoa boa e com força de vontade – que não é deprimida, apática ou está sofrendo de fraqueza de vontade - o que ela está motivada a fazer terá sido alterado em resposta à mudança em seu juízo moral, que é exatamente o que o internismo nos levaria a esperar. Se o internismo for verdadeiro, então, podemos prontamente explicar as mudanças motivacionais. A conexão confiável entre juízo moral e motivação é, em última análise, melhor explicada internamente como devido ao próprio conteúdo ou natureza do próprio juízo moral (SMITH, 1994, p. 72). É claro que, em última análise, aqueles que aceitam o internismo devem uma teoria da natureza dos juízos morais que explique e capture

a conexão necessária que supostamente existe entre o juízo moral e a motivação.[46]

Como tese diretamente oposta ao internismo motivacional, o **externismo motivacional** nega que a tese de que há uma conexão necessária entre juízo moral e motivação seja uma platitude conceitual. De acordo com o externismo, qualquer conexão que exista entre o juízo moral e a motivação é puramente contingente, mesmo que a explicação acabe se apoiando em características profundas da natureza humana.[47] A motivação moral ocorre quando um juízo moral se combina com um desejo, e o conteúdo do juízo está relacionado ao conteúdo do desejo de forma a racionalizar a ação. O argumento anterior em favor do internismo na verdade nega que o externismo possa explicar adequadamente o fenômeno básico da motivação moral e, em particular, a aparentemente confiável mudança de motivação moral para corresponder às mudanças de juízo moral. Mas, por que pensar que o externismo é explanatoriamente inadequado? Uma vez que temos a tese internista sobre a conexão necessária entre juízo moral e motivação, parece que temos, por assim dizer, a história completa: se um indivíduo faz um juízo moral, ele está, *ceteris paribus*, motivado; se ele não está motivado, então ele não está emitindo um juízo moral sincero e competente, não obstante as aparências em contrário. Como o externista nega a existência de uma conexão necessária entre o juízo moral e a motivação, a tese externista requer uma explicação independente da motivação moral. O internista sustenta que nenhuma explicação será suficiente - chamemos isso de “desafio internista” ao externismo.

De acordo com uma importante versão do desafio internista, oferecida por Michael Smith, o externista tem que explicar a conexão entre juízo moral e motivação **externamente** não como devida ao conteúdo dos juízos morais, mas, sim, ao "conteúdo das disposições motivacionais possuídas pela pessoa de boa e forte vontade.” (SMITH, 1994, p. 72). Mas isso supostamente compromete o externista

---

[46] Os não-cognitivistas consideram-se responsáveis por fornecer tal explicação, quando nos dizem que o conteúdo dos juízos morais é dado por um ou outro estado conativo. Para um esforço rigoroso de um cognitivista moral para oferecer uma explicação, ver Smith (1994). Para uma discussão crítica geral da teoria de Smith, veja Brink (1997), Copp (1997) e Sayre-McCord (1997), e veja a resposta de Smith (1997).

[47] Para uma discussão útil e informativa do debate internismo-externismo, vide Darwall (1983, cap. 5), Brink (1989, p. 37-80), Smith (1994, cap. 3), Svavarsdottir (1999) e Shafer-Landau (2003, cap. 6).

com uma imagem inaceitável da motivação moral. Segundo o internista, um agente que é movido a fazer a coisa certa é movido a fazer exatamente aquilo que é dado pelo conteúdo de seu juízo moral; ela está motivada a fazer exatamente o que julga ser correto (p. 73). A pessoa que julga, por exemplo, ser correto praticar um ato que promova o bem-estar de outra pessoa, adquire um desejo **não-derivado** de promover este bem-estar e por este desejo é movida. Em contraste, o externista precisa dizer que um agente é movido a fazer o que ele julga correto devido ao conteúdo das disposições motivacionais que ele tem ao ser uma boa pessoa. A questão então é quais podem ser essas disposições. Lembre-se de que quando o juízo de uma pessoa muda, sua motivação também tende a mudar. Se tal mudança motivacional deve ser explicada em termos das disposições motivacionais da pessoa boa, e não em termos do conteúdo de seus juízos morais, então a única disposição que poderia fornecer a explicação seria a motivação **para fazer a coisa certa**, seja o qual for o caso (p. 75). De acordo com Smith, o externismo implausivelmente trata a motivação moral como derivada; ou seja, ela deriva do desejo de fazer a coisa certa unido ao juízo moral do momento em que uma pessoa decide sobre a coisa certa a fazer. Não é o caso que uma pessoa deseja não-**derivadamente** promover o bem de outra porque ela julga correto promover este bem e, portanto, deseja fazer exatamente isso. Ao contrário, ela deseja fazer o que é certo, e neste momento simplesmente é o caso que promover este bem é o correto. Mas a pessoa boa, afirma Smith, deseja a justiça, a igualdade e o bem-estar de seus entes queridos de forma **não-derivada**. Desejar não-derivadamente apenas fazer o que se acredita ser correto [no sentido abstrato], estar motivado dessa forma, e não [desejar] justiça, a igualdade e o bem-estar [no sentido material e diretamente em si mesmas], é “um fetiche ou vício moral” (p. 75). Smith sugere que, ao tomar a pessoa boa como motivada a fazer o que ela acredita ser moralmente certo [no sentido abstrato], seja o que for, a teoria externista “a aliena dos fins aos quais a moralidade propriamente visa” (p. 76).

Externistas respondem a este desafio apontando que o fato de uma boa pessoa estar motivada [derivadamente] a fazer o que ela julga correto não a impede de também estar motivada de forma não-derivada, por valorização direta, a promover, por exemplo, o bem-estar de seus entes queridos. Eles também argumentam que não há nada fetichista em supor que a pessoa boa está motivacionalmente disposta a fazer a coisa certa e que, em qualquer caso, explicações externistas alternativas de uma conexão estável entre julgamento moral e motivação estão disponíveis (COPP, 1997, p. 49-50). Por exemplo, um indivíduo pode simplesmente estar disposto a desejar não-derivadamente a fazer qualquer coisa que ele acredita ser correto

fazer ou qualquer coisa que ele julga ter valor, ao invés de estar motivada derivadamente a fazer a coisa certa, seja o que for (COPP, 1997, p. 50-51). Sigrun Svavarsdottir (1999) argumenta que Smith está enganado quando afirma que a única opção externista para explicar a conversão motivacional é apelar para o desejo derivado de fazer a coisa certa, mas ela pensa que algo próximo à visão que Smith rejeita fornece exatamente a correta teoria externista da motivação moral. Na opinião dela, devemos entender que a pessoa boa está interessada em fazer o que é moralmente valioso ou obrigatório, onde esse interesse deve ser entendido como englobando o que é honesto, razoável, imparcial, gentil, solícito, justo e assim por diante. O fato de a pessoa boa estar motivacionalmente disposta desta maneira, não significa, como Smith parece sugerir, que ela se preocupa apenas com uma coisa, a saber, fazer o que acredita ser moralmente correto. Nem significa que ela realiza um ato concebendo-o simplesmente como a coisa certa a fazer. Pelo contrário, é compatível com o cenário externista que a boa pessoa muitas vezes simplesmente responderá diretamente à necessidade de conforto ou alívio de outra pessoa. Além disso, uma visão externista que concebe a pessoa boa como motivada pelo desejo de ser moral não envolve a introdução de um pensamento estranho (ou alienante) - "é a coisa certa a fazer" - em sua consciência a fim de explicar a motivação moral. Em vez disso, tendo formado o juízo moral de que ela deve executar a ação φ, o desejo de ser moral desempenha, na pessoa boa, um papel em efetuar a "transição psicológica" a partir do juízo de que a ação φ é correta para o desejo de executar φ (SVAVARSDOTTIR, 1999, p. 201). Finalmente, Svavarsdottir argumenta que embora o desejo de fazer a ação moralmente correta, digamos, ajudar outra pessoa em necessidade, possa derivar inicialmente de um desejo [abstrato] de ser moral, o desejo material direto pode vir a operar independentemente do desejo abstrato, de modo que seu desejo de ajudar não é simplesmente um desejo instrumental (SVAVARSDOTTIR, 1999, p. 205–206, 213-214).

De fato, o ponto em que os externistas insistem é que algum estado conativo precisa estar em ação na passagem do ato de julgar que a ação φ é correta para a condição de estar motivado a executar φ. Afinal, o externista nos lembra, esse movimento não ocorre em **todos** os agentes morais; alguns deles julgarão φ correto sem com isso estarem motivados para φ. Os externistas normalmente consideram isso como um postulado a partir da observação do senso comum de que existe uma ampla variação no impacto que os juízos morais têm sobre os sentimentos, deliberações e ações das pessoas (SVAVARSDOTTIR, 1999, p. 161).

Os debates entre internistas e externistas geralmente se concentram na figura do "amoralista" - a pessoa que aparentemente faz juízos morais, embora permaneça totalmente imotivado para obedecê-los. Vários esforços foram feitos para responder ao problema do amoralista, e esses esforços levaram ao desenvolvimento de numerosas versões de internismo motivacional. Geralmente, os internistas insistem que o amoralista é uma impossibilidade conceitual. A estratégia internista padrão para lidar com o hipotético amoralista é identificar um conteúdo para juízos morais cujo resultado seria que nenhum agente (ou nenhum agente **racional**, pelo menos) poderia empregar conceitos morais com competência ao expressar um juízo moral sincero, enquanto permanece imotivado. Os internistas admitem que a motivação moral não precisa ser predominante; desejos concorrentes podem ser mais fortes e, portanto, podem suplantá-la. Eles também concedem que a motivação moral é vulnerável; uma pessoa pode julgar a ação φ correta, embora não consiga ser movida para φ, devido à depressão ou fraqueza de vontade. Entretanto, casos de irracionalidade à parte, a pessoa que "**parece**" estar fazendo um juízo moral, embora permaneça imotivada, realmente ou não tem competência dos conceitos morais ou não está falando sinceramente. Neste último caso, ela julga um ato "correto" apenas no sentido de "discurso indireto" (HARE, 1963), como quando o criminoso impenitente, buscando uma sentença menor, diz ao juiz, em tom de remorso, que ele sabe que o que fez foi "errado".

Os externistas, é claro, afirmam que o amoralista não é uma impossibilidade conceitual. Afinal, se podemos **conceber** amoralistas, como certamente podemos, então eles não são conceitualmente impossíveis (SHAFER-LANDAU, 2003, p. 146). Ao contrário do que os internistas afirmam, os indivíduos podem aplicar conceitos morais com sinceridade e competência, sem estarem motivados de nenhuma maneira específica. Embora alguns amoralistas possam usar termos morais apenas no sentido de discurso indireto, nem todos os casos de falha motivacional podem ser explicados como casos de irracionalidade, incompetência conceitual ou falta de sinceridade.

Nesse ponto da dialética, internistas e externistas tendem a produzir argumentos adicionais em um esforço para superar o que parece um impasse. Externistas, por exemplo, podem nos convidar a considerar casos em que uma pessoa julga φ correto, embora acreditando que seria de fato impossível ter sucesso em executar φ, ou casos em que ela pensa que executar φ interferiria de forma significativa em seu bem-estar ou iria impedi-la de obter algo que ela deseja profundamente. Não parece plausível que, em tais casos, uma pessoa possa julgar φ correto, sem com isso estar motivada para φ? De forma mais geral, os externistas

argumentam que os internistas não conseguem dar sentido ao desafiante histórico da moralidade - o cético que pergunta: "Por que ser moral?" Os externistas também podem oferecer argumentos com o objetivo de transferir o ônus da prova para os internistas a fim de fazer com que este mostre que o internismo é uma platitude conceitual (SVAVARSDOTTIR 1999, 2006).

A fim de abordar mais completamente a variedade de casos em que falha a conexão entre o juízo moral e a motivação, os internistas oferecem versões mais qualificadas de internismo. Essas versões mais qualificadas sustentam que a conexão necessária entre juízo moral e motivação se mantém apenas sob certas condições. Quanto a quais são essas condições, vários pontos de vista foram apresentados. Por exemplo, Smith (1994) apresentou uma versão de internismo, ou o que ele chama de "condicional de praticabilidade", sustentando que existe uma conexão necessária entre o juízo moral e a motivação, pelo menos na "pessoa de boa e forte vontade". Ele afirma que a conexão entre juízo moral e motivação se mantém na pessoa que é "agencialmente racional". Outros filósofos sugerem que isso ocorre na pessoa que é “psicologicamente normal” (DREIER, 1990) ou na pessoa que é “moralmente sensível” (BJORKLUND *et al*., 2012, p. 126-128).

Mas problemas permanecem. E quanto à pessoa que fez um certo juízo moral no passado e sempre foi movida a agir de acordo com esse juízo, mas que deixa de ser motivada, enquanto continua a sustentar o juízo? Talvez ela julgue que moralmente deve trabalhar ativamente para ajudar os doentes e pobres. Após 20 anos fazendo isso, ela conclui que já fez o suficiente e deixa de estar motivada para agir de acordo com seu juízo, mas continua a julgar que moralmente deve trabalhar ativamente para ajudar os doentes e pobres. Não parece plausível que ela seja competente em relação aos conceitos morais, que fale com sinceridade, que use termos morais em seu sentido comum, em vez de no discurso indireto? E mesmo assim, ela não precisa ser apática, deprimida ou mentalmente doente de alguma forma.

Casos como esses levaram alguns a se afastar da ideia de que a conexão necessária entre o juízo moral e a motivação vale para todo e cada juízo moral individual. Aqueles que avançam formas do que Bjorklund *et al*., de 2012 (vide BJORNSSON *et al*., 2015), chamam de “internalismo deferido” afirmam que a conexão necessária entre juízo moral e motivação pode ser mais atenuada.

Internismo Deferido:
Necessariamente, se uma pessoa julga que moralmente deve executar a ação φ, então ou ela está (pelo

> menos em algum grau) motivada para executar φ, ou alguns de seus juízos morais relevantemente relacionados são acompanhados de motivação. (BJORKLUND *et al*., 2012, p. 128)

Portanto, no exemplo acima, nossa assistente social que deixou de ser motivada, mas ainda julga que moralmente deveria executar φ, possuía alguns juízos morais relevantemente relacionados que eram acompanhados por motivação, a saber, seus próprios juízos anteriores. Outros sustentam que o amoralista só é inteligível em relação às condições em que a motivação moral se mantém normal [estável entre os membros de uma comunidade], e isso levou alguns a defender que os juízos morais relevantemente relacionados são aqueles comuns ao seio da comunidade do agente. Tresan (2009a, 2009b), por exemplo, considera a possibilidade de uma versão "comunitária" do internismo deferido, que considera as crenças de um indivíduo como morais apenas quando, na comunidade do agente, as crenças com esse conteúdo são acompanhadas de motivação (*vide* BLACKBURN, 2001, p. 63).

À medida que as versões de internismo se tornam cada vez mais qualificadas, podemos nos inclinar a perguntar se ainda há alguma distinção entre internismo e externismo. Podemos questionar se algum deles tem qualquer vantagem em explicar a estabilidade da motivação moral. E, se for esse o caso, a história mais simples sobre a conexão entre o juízo moral e motivação contada pelo externista parecerá mais atraente.

Os externistas afirmam que podem explicar completa e adequadamente a forte, mas, em última instância, contingente conexão entre o juízo moral e a motivação, oferecendo várias explicações de como os juízos morais motivam de maneira estável. Como vimos, Svavarsdottir busca explicar a motivação moral apelando para um estado conativo particular, a saber, o desejo de fazer o que é moralmente valioso ou necessário – em resumo, o desejo de ser moral. Peter Railton invoca a preocupação que as pessoas geralmente têm de poder justificar suas escolhas e conduta de um ponto de vista mais imparcial. Mas ele parece pensar que os motivos mais comuns das pessoas também desempenham um papel; pelo menos isso é sugerido quando ele observa que, se realmente queremos que as pessoas levem a moralidade a sério, "devemos nos perguntar como podemos mudar a maneira como vivemos para que nossa conduta moral seja mais regularmente racional, dados os fins que realmente temos" (RAILTON, 1986a, p. 203). De acordo com David Brink, o externismo torna a força motivacional de nossos juízos morais

"uma questão de fato psicológico contingente, dependendo tanto do conteúdo das crenças morais das pessoas quanto de suas atitudes e desejos" (BRINK, 1989, p. 49). Ainda assim, essas atitudes e desejos podem ser amplamente compartilhados e enraizados em características centrais da natureza humana. Suponha, como sustentou o filósofo David Hume, que a simpatia seja uma característica profunda e amplamente compartilhada da psicologia humana. Então, Brink observa, embora possa ser um fato contingente que a maioria das pessoas tem o desejo de cumprir o que elas acreditam que a moralidade exige, isto também é um fato profundo sobre eles. "A motivação moral, sob tal ponto de vista, pode ser generalizada e previsível, mesmo que não seja necessária, nem universal, nem suprema" (BRINK, 1989, p. 49; BOYD, 1998, seção 4.7).

Filósofos que endossam o externismo normalmente também endossam o humianismo, embora seja importante lembrar que muitos internistas, incluindo a maioria dos não-cognitivistas e expressivistas sobre o juízo moral, também aceitam a teoria humiana da motivação. De fato, alguns afirmam que a observação básica que apoia o externismo também dá suporte à teoria humiana: a ampla variação no impacto motivacional dos juízos morais sugere não apenas que eles motivam contingentemente, mas que o fazem por meio de algum estado conativo. Ainda assim, os externistas não precisam ser humianos. Shafer-Landau, que rejeita o humianismo e o internismo, afirma que as crenças morais são de fato intrinsecamente motivadoras - elas podem motivar por si mesmas. Mas contra o internismo, elas não motivam necessariamente. Crenças intrinsecamente motivadoras podem falhar em motivar sob condições de extrema exaustão, séria depressão ou impulsos opostos intensos (SHAFER-LANDAU, 2003, p. 147-148). O fato de Shafer-Landau tratar a falibilidade da motivação moral sob tais condições como suporte a uma forma de externismo, enquanto Smith a trata sob condições semelhantes como compatível com uma forma de internismo, sugere alguma discordância entre os filósofos quanto a exatamente quando uma teoria deve ser classificada como uma forma de internismo ou externismo.[48]

---

[48] *Vide* Brink (1997, p. 7) comparando sua forma de caracterizar o internismo com a caracterização de Smith.

4. Motivação Moral e Metaética

O pensamento filosófico sobre o fenômeno da motivação moral há muito coincide e influencia os esforços contínuos para abordar as questões fundamentais da ética. De especial importância tem sido o uso de teses sobre a natureza da motivação moral para apoiar o **antirrealismo** na ética - a visão de que, ao contrário do que afirmam os realistas morais, não há fatos morais, nem verdades sobre o que a moralidade exige, proíbe ou permite, exceto, talvez, em algum sentido minimalista. Já vimos um exemplo de como as ideias sobre motivação moral podem ter relação com visões metaéticas mais amplas na crítica de Mackie ao objetivismo ético. Conforme observado anteriormente, Mackie defende o **antirrealismo cognitivista**, uma forma de antirrealismo que associa o cognitivismo a uma teoria do erro. De acordo com o antirrealismo cognitivista, embora sentenças éticas expressem proposições sobre propriedades objetivamente prescritivas - aquelas com a propriedade "a-ser-perseguida" embutida - tais propriedades não existem; e devido a esta falha de pressuposto, estamos sistematicamente em erro com relação a nossos juízos morais.

O desenvolvimento de teorias metaéticas ao longo dos últimos oitenta anos talvez tenha sido influenciado mais profundamente pelo uso de certas teses sobre motivação moral para apoiar o **antirrealismo não-cognitivista**. O antirrealismo não-cognitivista, assim como o antirrealismo cognitivista, rejeita a existência de propriedades morais e fatos morais robustos. Mas, ao contrário da última visão, rejeita o cognitivismo em favor do não-cognitivismo, que, como tradicionalmente retratado, é a teoria de que os juízos morais expressam atitudes em vez de crenças ou proposições e que, consequentemente, não possuem condição de verdade.[49]

Shafer-Landau (2003) oferece uma formulação do que ele chama de **Argumento Não-cognitivista**, que torna explícito como teses acerca da motivação moral foram empregadas para apoiar o antirrealismo não-cognitivista:

---

[49] Várias qualificações para essa caracterização do antirrealismo não-cognitivista são necessárias, embora uma discussão mais extensa nos levasse além do escopo deste verbete. Ambas as versões clássica e contemporânea da tese permitem sentidos nos quais podemos falar legitimamente sobre fatos morais e verdades morais. Ainda assim, o antirrealismo não-cognitivista rejeita a ideia de que os juízos morais são representacionais. Confira, por exemplo, Stevenson (1944), Gibbard (1990) e Blackburn (1984).

1 Necessariamente, se um indivíduo julga sinceramente uma ação correta, então, este individuo está motivado, em algum grau, a agir de acordo com esse juízo. (Internismo Motivacional)
2 Quando tomadas por si mesmas, as crenças não motivam nem geram quaisquer estados motivacionalmente eficazes. (Humianismo Motivacional)
3 Portanto, juízos morais não são crenças (Não-cognitivismo Moral)[50]

Uma vez que o realismo moral incorpora a afirmação cognitivista de que os juízos morais são crenças, a conclusão do argumento não cognitivista implica que o realismo moral é falso.

Filósofos contemporâneos que buscam defender versões de realismo moral ou objetivismo precisam lidar com essa linha básica de argumentação, mesmo que nem sempre tenham se engajado nela explicitamente. O Argumento Não-cognitivista, portanto, nos fornece uma ferramenta útil para mapear posições concorrentes em metaética. Podemos categorizar negativamente as posições filosóficas cotejando-as em termos de quais premissas do argumento não-cognitivista eles aceitam ou rejeitam.[51] Alguns rejeitam a premissa 1, muitas vezes defendendo formas de **realismo moral naturalista** que abraçam o externismo (*vide* RAILTON, 1986; BOYD, 1988; BRINK, 1989). De acordo com estes realistas, as propriedades morais são um tipo de propriedade natural e os fatos morais são fatos naturais. Os juízos sobre esses fatos expressam proposições e, portanto, podem ser verdadeiros ou falsos, mas esses juízos não motivam necessariamente. Se nossos juízos morais nos motivam, isso é determinado por fatos psicológicos contingentes e nossas

---

[50] Shafer-Landau (2003, p. 121) reconhece que nenhum filósofo apresentou o argumento apenas desta forma, mas ele afirma (p. 121, n. 1) que o argumento representa com precisão as ideias que encontramos em Stevenson (1963), Aiken (1950), Hare (1952, p. 79-93), Nowell-Smith (1954, p. 36-43), Harman (1975) e Mackie (1977, p. 27-42). Segundo Shafer-Landau, Simon Blackburn (1984, p. 188-189) também emprega este argumento embora "com algumas reservas".

[51] Este mapa obviamente não visa esgotar a gama de pontos de vista positivos que os filósofos podem aceitar. Aceitar ou rejeitar premissas específicas do argumento não-cognitivista nem sempre nos compromete com uma teoria metaética positiva específica.

crenças morais substantivas. Alguns rejeitam a premissa 2 (MCDOWELL, 1978, 1979), e entre estes alguns rejeitam a premissa 2 alinhando-se a versões de construtivismo moral ou racionalismo (*vide* DARWALL, 1983; SCANLON, 1998). Estas últimas teorias assumem formas amplamente variadas, mas geralmente veem os princípios morais como princípios racionais, ou como o resultado de um acordo hipotético entre pessoas razoáveis e adequadamente situadas. Razões morais são considerações motivadoras, pelo menos quando refletimos sobre elas, mas sua força motivadora não depende de um desejo anterior. Alguns rejeitam as premissas 1 e 2, defendendo formas de **realismo moral não naturalista** (SHAFER-LANDAU, 2003). Segundo esta teoria, as propriedades morais não são idênticas às propriedades naturais ou descritivas, embora possam ser inteiramente constituídas por elas. Os juízos morais são intrinsecamente motivadores - eles podem motivar na ausência de um desejo preexistente, mas não são necessariamente motivadores. Finalmente, alguns aceitam ambas as premissas 1 e 2, pelo menos apropriadamente refinadas, argumentando que podemos entendê-las como compatíveis com o cognitivismo moral e, além disso, com o realismo moral (SMITH, 1994). Por exemplo, Smith entende os juízos normativos como representando razões normativas. Segundo ele, as razões normativas são dadas por fatos sobre o que nós, adequadamente idealizados, desejaríamos desejar; e a existência de tais fatos significa que alguns desejos são racionalmente exigidos. Se acreditamos que temos uma razão normativa para φ, então, racionalmente, devemos executar φ, e ao julgar que temos uma razão normativa para φ, estaremos necessariamente motivados para φ, na medida em que formos racionais. O conceito de correção é o conceito do que desejaríamos desejar em nosso mundo real, se fôssemos totalmente racionais. Quando acreditarmos que seria certo φ, então, estaremos, na medida em que formos racionais, motivados para φ.

Mais recentemente, alguns filósofos (*vide* TRESAN, 2006, 2009a, 2009b) argumentam que o internismo motivacional, quando entendido como o que Bjorklund *et al*. (2012) chamam de **internismo motivacional não-constitucional**, é compatível não apenas com o racionalismo smithiano, mas com uma ampla variedade de teorias metaéticas cognitivistas. Internalismo não-constitucional (ou o que Tresan chama de "internismo SM") é a visão de que, de acordo com nosso conceito de uma crença moral, um estado mental conta como uma crença moral apenas se for acompanhado por motivação. Necessariamente, se p é uma crença moral, então p é acompanhado de motivação. De acordo com Tresan, uma vez que reconhecemos essa forma de internismo, vemos que ela é compatível com quase qualquer versão de cognitivismo e, portanto, com uma gama de visões metaéticas, incluindo formas de naturalismo

moral (2006, p. 68). Isso contrasta com as tentativas de combinar internismo e cognitivismo com base na tese de que a natureza da crença moral é tal que garante a motivação (pelo menos sob certas condições) seja por causa do conteúdo das crenças morais (SMITH, 1994) ou porque as crenças morais são elas mesmas intrinsecamente estados motivadores (DANCY, 1993). (*vide* BJORKLUND *et al*., 2011).

5. Motivação Moral e Psicologia Experimental

Pesquisas em psicologia experimental também contribuem para a forma como entendemos e respondemos às nossas perguntas sobre motivação moral. Recentemente vários filósofos têm recorrido à psicologia para lidar com questões de metaética e, em particular, com a questão da motivação moral. Argumenta-se que avanços em psicologia podem apontar implicações para a natureza da motivação em geral, para o debate entre internistas e externistas motivacionais e para a plausibilidade de várias teorias filosóficas acerca da natureza da motivação moral.

Schroeder et al. esboçam o que eles descrevem como caricaturas de quatro possíveis teorias da motivação moral, que eles rotulam de instrumentalista, cognitivista, sentimentalista e personalista, e argumentam que "as teorias da motivação moral que melhor se adaptam ao quadro científico atual são aquelas que devem muito mais a Hume ou Aristóteles do que Kant "(72).[52] De acordo com o instrumentalista, "as pessoas são motivadas quando formam crenças sobre como satisfazer desejos [intrínsecos] preexistentes" (74), o que leva, por sua vez, à formação de desejos não-intrínsecos de realizar ações específicas destinadas a satisfazer seus desejos intrínsecos. Quando uma pessoa tem um desejo intrínseco, D, e passa a acreditar que a ação φ satisfará D, ela passa a desejar (não-intrinsecamente) executar φ. Na visão instrumentalista, comumente chamada de "humiana", o caráter especificamente moral da motivação vem de um desejo intrínseco de fazer o que se julga correto. Em contraste com o instrumentalista, o cognitivista sustenta que a motivação moral

---

[52] Embora ofereçam as quatro teorias possíveis da motivação moral como caricaturas, Schroeder et al. identificam pontos de vista na literatura com os quais algumas dessas caricaturas se assemelham: instrumentalista (WILLIAMS, 1981), cognitivista (KORSGAARD, 1994; MCDOWELL, 1998, cap. 4; SMITH 1994) e personalista (ARISTOTLE, 2000; HURSTHOUSE, 1999; SLOTE, 2001).

começa, não com desejos, mas com crenças sobre quais ações são corretas. Essas crenças motivam independentemente de desejos intrínsecos preexistentes. O caráter especificamente moral da ação surge não a partir de desejos, pelo menos não em primeira instância, mas de juízos morais (p. 76). O sentimentalista vê as emoções como desempenhando um papel central na motivação moral e, para que uma ação resulte desta motivação, certas emoções devem causar essa ação. Compaixão ou simpatia são exemplos destas emoções (p. 77). Finalmente, o personalista vê a fonte da motivação moral no caráter moralmente bom, mais especificamente, nas virtudes. “O bom caráter envolve o conhecimento do bem, desejar o bem em si mesmo, disposições emocionais duradouras que favorecem a boa ação e hábitos duradouros de responder com boas ações aos próprios conhecimentos, desejos e emoções” (77). A ação moral começa quando a sensibilidade de uma pessoa aos padrões morais e heurísticas morais (como mentir tende a ser ruim) a leva a julgar que uma ação seria boa, o que por sua vez ativa suas disposições e hábitos emocionais duradouros, resultando assim em motivação moral.

De acordo com Schroeder *et al*., cada uma dessas teorias “pressupõe compromissos relativos à natureza dos estados psicológicos, como crenças, desejos, escolhas, emoções e assim por diante, juntamente com compromissos relativos aos papéis funcionais e causais que desempenham” (p. 79). Como esses compromissos não são apenas filosóficos, mas também empíricos, estes filósofos buscam selecionar contribuições empíricas – “o livro didático da neurociência” – sobre a neurofisiologia da motivação e avaliam as implicações destas pesquisas para aquelas quatro teorias filosóficas sobre a motivação moral.

Schroeder *et al*. argumentam que a visão instrumentalista se sai bem com relação a neurociência, assim como o personalismo. Em contraste, a explicação cognitivista da motivação moral enfrenta problemas “uma vez que nosso comportamento moral não parece estar sob o controle de estados cognitivos apenas, independentemente dos desejos” (p. 106). O sentimentalismo também enfrenta dificuldades "porque o sistema emocional, embora intimamente ligado ao sistema subjacente à ação voluntária, acabará sendo, no entanto, distinto dele, a menos que as próprias emoções sejam construídas em parte a partir dos desejos" (p. 106). Schroeder et al. reconhecem que nossa compreensão atual da neurociência é incompleta e que respostas podem ser oferecidas às críticas que eles fazem (p. 106). No entanto, eles sugerem que a atenção à neurociência “poderia servir para restringir futuras teorias sobre a estrutura da agência moral [...]” (p. 107).

Roskies (2003) busca extrair conclusões sobre um tipo particular de internismo motivacional, concentrando-se em evidências empíricas relacionadas a pacientes com danos no córtex ventromedial (VM). Mais precisamente, ela argumenta contra a visão de que a crença moral envolve motivação moral, uma visão que ela chama de "internismo de motivos", que é a visão chamada de "internismo motivacional forte" na seção 3.2. Conforme ela descreve, internismo de motivos é a teoria segundo a qual "a motivação é intrínseca ou um componente necessário da crença ou juízo moral" (52). A pessoa que sinceramente acredita que deve executar φ está motivada para tanto, pelo menos até certo ponto. Em contraste, o externista sustenta que a crença moral não acarreta motivação moral; uma pessoa pode acreditar que deve executar φ, embora não tenha qualquer motivação para φ. Roskies explica que a afirmação internista envolve necessidade, intrinsecalidade e especificidade. De acordo com o internismo de motivos, é uma verdade **necessária** que a motivação acompanhe a crença ou juízo moral e, portanto, é uma verdade "para qualquer agente capaz de crenças morais" (p. 52). A **intrinsecalidade** do internismo de motivos consiste na ideia de que a conexão entre crença ou juízo moral e motivação se mantém por causa do conteúdo da crença moral, e não em função de algo não relacionado ao conteúdo dessa crença. Quanto à **especificidade**, o internismo de motivos vê as crenças morais como diferentes de outros tipos de crenças, que não são intrinsecamente motivadoras (52).

De acordo com Roskies, o internismo de motivos enfrenta um dilema: ou sua tese internista é muito fraca e, portanto, filosoficamente desinteressante, ou é forte o suficiente para ser filosoficamente interessante, mas também "demonstravelmente falsa" (p. 51). No primeiro vetor do dilema, a tese internista é muito fraca e, portanto, filosoficamente desinteressante. Roskies o ilustra com a visão de Smith de que há uma conexão necessária entre o juízo moral e a motivação, exceto quando um agente deixa de ser agencialmente racional. Esta tese requer uma especificação do que é ser agencialmente racional, mas Roskies argumenta que se esta condição equivale a desejar agir como se julga melhor, então a tese é trivial. Não é uma afirmação forte sobre uma conexão necessária entre o juízo moral e a motivação, mas apenas aponta um elemento da definição de racionalidade agencial. Além disso, a tese carece de especificidade, pois também se aplica ao que um agente julga ser melhor no sentido não-moral. Outras versões do internismo de motivos tornam a tese igualmente trivial (p. 53-55).

No outro vetor do dilema, a afirmação internista é filosoficamente interessante, porém falsa. Aqui, Roskies argumenta que as pessoas com lesão no córtex VM são

um “contraexemplo ambulante” ao internismo. Conforme Roskies os descreve, tais pacientes “parecem cognitivamente normais em um amplo espectro de testes psicológicos padrão, incluindo aqueles que medem a inteligência e as habilidades de raciocínio”, no entanto, eles “todos parecem ter particular dificuldade em agir de acordo com os costumes sociais, apesar de sua continuada capacidade para julgar apropriadamente em tais situações” (p. 56). A condição desses pacientes tem sido chamada de “sociopatia adquirida”. De acordo com Roskies, os pacientes VM são capazes de "raciocinar moralmente em um nível normal" e seus juízos morais "estão de acordo com aqueles dos normais", entretanto eles falham em agir de forma confiável como as pessoas normais fazem, assim como "parecem carecer de motivação e respostas emocionais adequadas” (p. 57). Os pacientes VM não exibem resposta da atividade eletrodérmica (RAE) a estímulos com carga emocional que as pessoas normais exibem, o que Roskies considera uma evidência da ausência de motivação. Os pacientes VM, supostamente, apresentam um contraexemplo para o internismo de motivos, já que eles possuem domínio dos termos morais e parecem fazer juízos morais sinceros, embora não tenham qualquer motivação para agir de acordo com eles (p. 59).

Vários argumentos contra os alegados resultados de pesquisas empíricas têm sido apresentados em favor do internismo motivacional. Alguns argumentam que os pacientes VM não possuem o domínio de conceitos morais (KENNETT; FINE, 2007), que os pacientes VM expressam juízos morais apenas no sentido que R. M Hare (1963) chamou de "discurso indireto" (KENNETT; FINE, 2007), que os pacientes VM estão de fato motivados quando emitem juízos morais (KENNETT; FINE, 2007), que a evidência oferecida para se alegar que os pacientes VM expressam juízos morais sem sentir emoções "não é decisiva" (PRINZ, 2015), que temos motivos para questionar se pacientes VM sustentam crenças morais (CHOLBI, 2006) e que os pacientes VM têm agência prejudicada e, portanto, não fazem julgamentos morais genuínos, que por sua vez exigem agência (GERRANS; KENNETT, 2010). De várias maneiras, essas alegações questionam se seria conceitualmente coerente tratar casos de pacientes VM como casos de amoralismo. Na medida em que a discordância diz respeito à coerência conceitual do amoralismo, é incerto como o apelo à literatura empírica ajuda a avançar o debate. Claro, Roskies pode, seguindo Prinz (2015), sustentar que o internismo é na verdade uma tese psicológica e não conceitual, caso em que essas críticas à coerência conceitual de tratar casos de pacientes VM como casos de amoralismo não se aplicariam.

Independentemente de as respostas de Roskies (2006, 2007), a muitas dessas objeções, serem ou não bem-sucedidas, há razões para duvidar de que os dados que Roskies cita sejam suficientes para minar o internismo. A própria Roskies reconhece que algumas versões de internismo (embora sejam algumas que ela considere problemáticas ou ainda insuficientemente desenvolvidas) podem ser consistentes com os dados sobre pacientes VM (2003, p. 62-63) Schroeder *et al*. (2010, p. 95) observam que pesquisas sugerem que os psicopatas têm uma "capacidade diminuída de distinguir violações morais das convencionais", o que levou alguns a concluir que eles possuem "conceitos morais ineptos" (NICHOLS, 2004 *apud* SCHROEDER *et al*., 2010, p. 96). Se eles realmente operam conceitos morais ineptos, então não representam problema para o internista. No caso de pacientes VM, no entanto, eles observam que "tem sido argumentado que as pessoas que exibem sociopatia adquirida não apresentam déficits morais, mas que seus déficits em aspectos não-morais da vida apenas se manifestam ocasionalmente em situações morais". Porém eles indicam que serão necessários mais estudos para resolver a questão sobre se os pacientes VM representam uma ameaça para o que Schroeder et al. chamam de cognitivismo e Roskies chama de internismo de motivo (2010, p. 97). Em qualquer caso, é questionado qual a melhor forma de explicar os dados existentes sobre pacientes VM. Os pacientes VM, que sofrem lesões no início da vida, exibem comportamento sociopático, incluindo comportamento violento, enquanto os pacientes VM que adquirem sua condição mais tarde na vida não apresentam os mesmos comportamentos. A questão em disputa é se os últimos pacientes não são violentos, por exemplo, porque seus juízos morais são motivadores em algum grau ou se seu comportamento não-violento é uma função do hábito (p. 98). Finalmente, como Roskies (2007, p. 205) explica cuidadosamente, "a evidência é inconclusiva porque ainda não se realizaram testes de cognição e comportamento de pacientes VM mais bem elaborados".

Em contraste com Roskies, Prinz (2015) argumenta que a evidência empírica sustenta o internismo. Ele argumenta primeiro que "o internismo pode ser entendido como uma tese psicológica" (p. 61), ao invés de uma verdade conceitual ou a priori, e em seguida oferece vários argumentos empíricos para apoiá-la. O primeiro invoca uma teoria denominada **sentimentalismo**.

1 Os juízos morais consistem em atitudes emocionais.
2 As atitudes emocionais são motivadoras.
3 Portanto, os juízos morais são motivadores. (p. 70)

O passo polêmico do argumento é a primeira premissa. Prinz diz que a premissa 1 é a própria tese do sentimentalismo, a visão de que "os juízos morais consistem em sentimentos dirigidos a tudo o que moralizamos" (p. 70). Ele argumenta que a tese apoia várias previsões empíricas, que estudos confirmam. Em primeiro lugar, estudos de neuroimagem sobre cognição moral fornecem evidências de que "as pessoas entram em estados emocionais quando fazem juízos morais" (p. 71). Em segundo lugar, estudos mostram que emoções induzidas têm impacto no juízo moral, e "emoções diferentes têm efeitos diferentes" (p. 72). Por exemplo, induzir nojo leva as pessoas a julgar um cenário envolvendo incorreção moral de forma mais severa. A felicidade induzida aumenta os juízos morais positivos, mas não os negativos, enquanto a raiva aumenta os juízos morais negativos, mas não os juízos morais positivos (p. 72). Finalmente, pessoas com disposições emocionais diferentes diferem em juízos morais. Por exemplo, "psicopatas, que apresentam déficits em várias emoções negativas, mas não nojo, mostram insensibilidade a crimes contra pessoas, porém não são conhecidos por desvios sexuais" (p. 73), enquanto aqueles com doença de Huntington apresentam tanto déficits em nojo como também exibem padrões de desvio sexual (p. 72). Prinz afirma que essas descobertas "adicionam suporte à tese de que as emoções são componentes dos juízos morais. As emoções ocorrem quando as pessoas expressam juízos morais, as emoções são usadas como informação nos relatos acerca da força das atitudes morais, e os déficits emocionais levam a déficits correspondentes na sensibilidade moral" (p. 73). De acordo com Prinz, dadas essas evidências, e dadas décadas de pesquisas que vinculam as emoções ao comportamento, apoiando assim a premissa 2, esse argumento fornece um forte suporte para o internismo.

Entretanto, a tese sentimentalista é muito menos clara do que Prinz deseja. Às vezes, ele diz que os juízos morais "consistem em" atitudes emocionais, às vezes que as emoções são "componentes" dos juízos morais. No entanto, não está claro que as evidências fornecem suporte adequado para o sentimentalismo, em oposição à tese de que as emoções (contingentemente) acompanham o juízo moral. Que emoções acompanhariam o juízo moral não é surpreendente, dada a importância da moralidade para o bem-estar humano.

Prinz oferece quatro argumentos adicionais, apenas dois dos quais são considerados brevemente aqui. O primeiro é um argumento a partir de evidências experimentais que parecem mostrar que as pessoas normalmente "consideram as

emoções como necessárias para ter atitudes morais" (p. 75). Por exemplo, os participantes de um estudo foram solicitados a avaliar as atitudes morais de dois indivíduos. A pessoa A fuma maconha, não se sente culpada por isso, não tem atitudes negativas para com os que o fazem, mas, mesmo assim diz que acha que fumar maconha é moralmente errado. A pessoa B fuma maconha, sente-se culpada por isso, tem sentimentos negativos em relação aos que fumam, mas, mesmo assim diz que acha que fumar maconha não é moralmente errado. A maioria dos entrevistados concluiu que o primeiro fumante não acredita que fumar maconha é moralmente errado, enquanto o segundo realmente acha que é moralmente errado, apesar de suas alegações em contrário. Prinz diz que "a maioria trata as emoções como necessárias e suficientes para atribuição moral", o que sugere que "a população comum está comprometida com uma espécie de sentimentalismo". Isso, no máximo, nos mostraria o que as pessoas comuns pensam; não apoiaria o sentimentalismo como uma tese metafísica "de que faz parte da essência dos juízos morais que eles podem ser motivadores" (p. 64). Prinz especula que mais pesquisas podem "estabelecer uma ligação conceitual mais consistente entre emoções e juízos morais" (p. 76). No momento, porém, os estudos que Prinz cita são insuficientes para dar muito apoio ao internismo.

O segundo argumento de Prinz invoca estudos que mostram que as pessoas às vezes são motivadas a agir moralmente, mesmo quando seria instrumentalmente racional não o fazer. Ele sugere que a "explicação mais simples é que os juízos morais têm força motivacional, independente da motivação não-moral" (p. 77). Mas os externistas estão pelo menos bem posicionados para explicar as evidências apelando para a aculturação e desejos comuns de ajudar os outros ou de fazer a coisa certa.

A pesquisa empírica sobre motivação moral é, naturalmente, de grande interesse por si só. Mas muitos duvidam da relevância de tal pesquisa para o que os internistas consideram uma tese conceitual, uma afirmação sobre uma conexão necessária entre o juízo moral e a motivação moral. Mesmo supondo que a pesquisa empírica pudesse ajudar a resolver o impasse entre internistas e externistas, a pesquisa existente está longe de fazê-lo.

Referência bibliográfica

AIKEN, H., Evaluation and Obligation, **Journal of Philosophy**, 47: 5–22, 1950.

ALSTON, W., Motives and Motivation, in P. Edwards (ed.), **The Encyclopedia of Philosophy**, New York: Macmillan, 399–409,1967.

ANSCOMBE, G. E. M., **Intention**, 2nd edition, Ithaca: NY: Cornell University Press,1963.

AYER, A.J., **Language, Truth, and Logic**, London: Gollancz, 1936,

BEDKE, M.S., Moral judgment purposivism: saving internalism from amoralism, **Philosophical Studies**, 144: 189–209, 2009.

BJORKLUND, F., Bjornsson, G., Erikson, J., Olinder, R., and Strandberg, C., Recent Work on Motivational Internalism, **Analysis**, 72: 124–137, 2012.

BJORNSSON, G., How emotivism survives immoralists, irrationality, and depression, **Southern Journal of Philosophy**, 40: 327–344, 2002.

BJORNSSON, G., Strandberg, C., Ollinder, R., Eriksson, J, and Bjorklund, F. (ed.), **Motivational Internalism**, Oxford: Oxford University Press, 1–20, 2015.

BLACKBURN, S., **Spreading the Word**, Oxford: Oxford University Press, 1984.

BLACKBURN, S., Errors and the Phenomenology of Value, in T. Honderich (ed.), **Morality and Objectivity: A Tribute to John Mackie**, London: Routledge & Kegan Paul, 1985.

BLACKBURN, S., How to Be an Ethical Anti-Realist, in **Essays in Quasi-Realism**, Oxford: Oxford University Press, 1987.

BLACKBURN, S., **Ruling Passions**, Oxford: Clarendon Press, 2001.

BOYD, R., How to be a Moral Realis, in G. Sayre-McCord (ed.), **Essays on Moral Realism**, Ithaca, NY: Cornell University Press, 181–228, 1988.

BRINK, D., **Moral Realism and the Foundations of Ethics**, Cambridge: Cambridge University Press, 1989.

BRINK, D., Moral Motivation**, Ethics**, 108: 4–32, 1997.

BROMWICH, D., Clearing conceptual space for cognitivist motivational internalism**, Philosophical Studies**, 148: 343–367, 2010.

BROOME, H., Reason and Motivation, **Proceedings of the Aristotelian Society** (Supplement), 71: 131–147, 1997.

CHOLBI, M. J., Belief attribution and the falsification of motive internalism, **Philosophical Psychology**, 19: 607–616, 2006.

CLARK, P., Kantian Morals and Humean Motives, **Philosophy and Phenomenological Research**, 68: 109–126, 2004.

COPP, D., Moral Obligation and Moral Motivation, **Canadian Journal of Philosophy**, Supplement, 21: 187–219, 1995.

COPP, D., Belief, Reason, and Motivation: Michael Smith's The Moral Problem, **Ethics**, 108: 33–54, 1997.

COPP, D. and SOBEL, D., Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology, **Social Theory and Practice**, 28: 243–276, 2002.

DANCY, J., **Moral Reasons**, Oxford: Basil Blackwell, 1993.

DARWALL, S., **Impartial Reason**, Ithaca, NY: Cornell University Press, 1983.

DEIGH, J., Empathy and Universalizability, **Ethics**, 105: 743–763, 1995.

DRIER, J., Internalism and Speaker Relativism, **Ethics**, 101: 6–26, 1990.

DRIER, J., Dispositions and Fetishes: Externalist Models of Moral Motivation, **Philosophy and Phenomenological Research**, 61: 619–638, 2000.

FRANKENA, W., Obligation and Motivation in Recent Moral Philosophy, in K. Goodpaster (ed.), **Perspectives on Morality: Essays of William Frankena**, Notre Dame, IN: Notre Dame University Press, 1976.

GERARD, E., and McNaughton, D., Mapping Moral Motivation, **Ethical Theory and Moral Practice**, 1: 45–59, 1998.

GIBBARD, A., **Wise Choices, Apt Feelings**, Cambridge, Mass.: Harvard University Press, 1990.

Greenberg, M., Moral Concepts and Moral Motivation, **Philosophical Perspectives**, 23: 137–164, 2009.

HARE, R., **The Language of Morals**, Oxford: Oxford University Press, 1952.

HARE, R., **Freedom and Reason**, Oxford: Oxford University Press, 1963.

HARMAN, G., Moral Relativism Defended, **Philosophical Review**, 85: 3–22, 1975.

HERMAN, B., **The Practice of Moral Judgement**, Cambridge, MA: Harvard University Press, 1993.

HUME, D**., An Enquiry Concerning the Principles of Morals**, Oxford: Clarendon Press, 1975.

HUME, D**., A Treatise of Human Nature**, Oxford: Clarendon Press, 1978.

JACKSON, F. and Pettit, P., Moral Functionalism and Moral Motivation, **Philosophical Quarterly**, 405: 20–40, 1995.

KAUPPINEN, A., Moral internalism and the brain, **Social Theory and Practice**, 34: 1–24, 2008.

KENNETT, J., What's Required for Motivation by Principle? in G. Bjornsson, C. Strandberg, R. Ollinder, J. Eriksson, and F. Bjorklund (eds.), **Motivational Internalism**, Oxford: Oxford University Press, 2015.

KENNETT, J. and Fine, C., Internalism and the Evidence from Psychopaths and Acquired Sociopaths, in W. Sinnott-Armstrong (ed**.), Moral Psychology, Vol. 3: The Neuroscience of Morality**, Cambridge, MA: MIT Press 173-190, 2008.

KORSGAARD, C., **The Sources of Normativity**, Cambridge: Cambridge University Press, 1996.

KRISTJANSSON, K., Aristotelian motivational externalism, **Philosophical Studies**, 164: 419–442, 2013.

KRIEGEL, U., Moral Motivation, Moral Phenomenology, and the Alief/Belief Distinction, **Australasian Journal of Philosophy**, 90: 469–486, 2012.

LITTLE, M., Virtue as Knowledge: Objections from the Philosophy of Mind, **Noûs**, 31: 59–79, 1997.

MACKIE, J., **Ethics: Inventing Right and Wrong**, New York: Penguin, 1977.

MAMELI, M., Evolution, Motivation, and Moral Beliefs, in K. Sterelny, R. Joyce, B. Calcott, and B. Fraser, (eds.), **Cooperation and its Evolution**, Cambridge, MA: MIT Press, 528–548, 2013.

MASON, E., An Argument Against Motivational Internalism, **Proceedings of the Aristotelian Society** (New Series), 108: 135–156, 2008.

MCDOWELL, J., Are Moral Requirements Hypothetical Imperatives? **Proceedings of the Aristotelian Society** 52: 13–42, 1978.

MCDOWELL, J., Virtue and Reason, **Monist**, 62: 331–50, 1979.

MCNAUGHTON, D., **Moral Vision**, Oxford: Basil Blackwell, 1988.

MILL, J., **Utilitarianism**, Indianapolis: Hackett Publishers, 1979.

MILLER, C., "Motivation in Agents", **Noûs**, 42: 222–266, 2008.

MILLER, C., Motivational Internalism, **Philosophical Studies**, 139: 233–255, 2008.

NAGEL, T., **The Possibility of Altruism**, Oxford: Oxford University Press, 1970.

NICHOLS, S., How psychopaths threaten moral rationalism: is it irrational to be amoral? **Monist,** 85: 285–304, 2002.

NOWELL-SMITH, P.H., **Ethics**, London: Penguin Books, 1954.

PARFIT, D., Reasons and Motivation, **Proceedings of the Aristotelian Society** (Supplement), 71: 99–130, 1997.

PLATTS, M., **Ways of Meaning**, London: Routledge and Kegan Paul, 1979.

PLATTS, M., Moral Reality and the End of Desire, in M. Platts (ed.), **Reference, Truth, and Reality**, London: Routledge and Kegan Paul, 1980.

PRINZ, J., An Empirical Case for Motivational Internalism, in G. Bjornsson, C. Strandberg, R. Ollinder, J. Eriksson, and F. Bjorklund (eds.), **Motivational Internalism**, Oxford: Oxford University Press, 61–84, 2015.

RAILTON, P., Moral Realism, **Philosophical Review**, 95: 163–207, 1986a.

RAILTON, P., Facts and Values, **Philosophical Topics**, 14: 5–31, 1986b.

RAILTON, P., Some Questions About the Justification of Morality, **Philosophical Perspectives** 6: 27–53, 1992.

ROSKIES, A., Are ethical judgments intrinsically motivational? Lessons from 'acquired sociopathy'**, Philosophical Psychology**, 16: 51–66, 2003.

ROSKIES, A., Patients with ventromedial frontal damage have moral beliefs, **Philosophical Psychology**, 19: 617–627, 2006.

ROSKIES, A., Internalism and the evidence from pathology, in W. Sinnott-Armstrong (ed.), **Moral Psychology, Vol. 3: The Neuroscience of Morality,** Cambridge, MA: MIT Press, 191-206, 2007.

ROSS, W.D., **The Right and the Good**, Oxford: Clarendon Press, 1930.

SAYRE-MCCORD, G., ´The Metaethical Problem, **Ethics**, 108: 55–83, 1997.

SCANLON, T., **What We Owe to Each Other**, Cambridge, Mass.: Harvard University Press, 1998.

SCHEULER, G., **Desire**, Cambridge, MA: MIT Press, 1995.

SCHROEDER, T., Roskies, A., and Nichols, S., Moral Motivation, in J. Doris (ed.), **The Moral Psychology Handbook**, Oxford: Oxford University Press, 72–110, 2010.

SHAFER-LANDAU, R., Moral Motivation and Moral Judgment, **Philosophical Quarterly**, 48: 353–8, 1998.

SHAFER-LANDAU, R., A Defence of Motivational Externalism, **Philosophical Studies** 97: 267–91, 2000.

SHAFER-LANDAU, R., **Moral Realism: A Defence**, Oxford: Clarendon Press, 2003.

SHAVER, R., Sidgwick on Moral Motivation, **Philosophers Imprint**, 6(1), 2006.

SHOEMAKER, D., Reductionist Contractualism: Moral Motivation and the Expanding Self**, Canadian Journal of Philosophy**, 30: 343–370, 2000.

SMIT, H., Internalism and the Origin of Rational Motivation, **Journal of Ethics**, 7: 183–231, 2003.

SMITH, M., The Humean Theory of Motivation, **Mind**: 36–61, 1987.

SMITH, M., **The Moral Problem**, Oxford: Basil Blackwell, 1994.

SMITH, M., Internalism's Wheel, in B. Hooker (ed.), **Truth in Ethics**, Oxford: Blackwell, 69–94, 1996.

SMITH, M., In Defense of The Moral Problem: A Reply to Brink, Copp, and Sayre-McCord, **Ethics**, 108: 84–119, 1997.

STAMPE, D., The Authority of Desire, **Philosophical Review**, 96: 335–81, 1987.

STEVENSON, C., **Ethics and Language**, New Haven: Yale University Press, 1944.

STEVENSON, C., The Emotive Meaning of Ethical Terms, in **Facts and Values**, New Haven: Yale University Press, 1963.

STRANDBERG, C., The Pragmatics of Moral Motivation, **Journal of Ethics**, 15: 341–369, 2001.

SVAVARSDOTTIR, S., Moral Cognitivism and Motivation, **Philosophical Review**, 108: 161–219, 1999.

SVAVARSDOTTIR, S., How Do Moral Judgments Motivate? in J. Dreier (ed.), **Contemporary Debates in Moral Theory**, Malden, MA: Blackwell, 163–181, 2006.

TIFFANY, E., Alienation and Internal Reasons for Action, **Social Theory and Practice**, 29: 387–418, 2003a.

TIFFANY, E., A Functional Account of Moral Motivation, **Southern Journal of Philosophy**, XLI: 601–625, 2003b.

TIMMONS, M., **Morality Without Foundations**, Oxford: Oxford University Press, 1999.

TRESAN, J., De dicto internalist cognitivism, **Noûs**, 40: 143–165, 2006.

TRESAN, J., Metaethical Internalism: Another Neglected Distinction, **Journal of Ethics**, 13: 51–72, 2009a.

TRESAN, J., The challenge of communal internalism, **The Journal of Value Inquiry**, 43: 179–199, 2009b.

VAN ROOJEN, M., Moral Rationalism and Rational Amoralism, **Ethics**, 120: 499–525, 2010.

VELLEMAN, J. D., The Guise of the Good, **Noûs**, 26: 3–26, 1992.

VELLEMAN, J. D., The Possibility of Practical Reason, **Ethics**, 106: 694–726, 1996.

WALLACE, R.J., Moral Motivation, in J. Dreier (ed.), **Contemporary Debates in Moral Theory**, Malden, MA: Blackwell, 182–196, 2006.

WEDGWOOD, R., The Metaethicists' Mistake, **Philosophical Perspectives**, 18: 405–426, 2004.

WIGGINS, D., Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs, **Proceedings of the Aristotelian Society** (New Series), 91: 61–85, 1990–1991.

WILLIAMS, B., Internal and External Reasons, in **Moral Luck**, Cambridge: Cambridge University Press, 1981.

WONG, D., Emotion and the Cognition of Reasons in Moral Motivation**, Philosophical Issues**, 19: 343–367, 2009.

ZANGWILL, N., Externalist Moral Motivation, **American Philosophical Quarterly**, 40: 143–154, 2003.

(IV) Sentimentalismo Moral*

Autoria: Antti Kauppinen
Tradução: Sérgio Farias de Souza Filho
Revisão: Rafael Cavalcanti de Souza

Para os sentimentalistas morais, nossas emoções e desejos desempenham um papel proeminente na anatomia da moralidade. Alguns creem que pensamentos morais são fundamentalmente sentimentais, outros que fatos morais são relacionados às nossas respostas sentimentais, ou que as emoções são as fontes primárias do nosso conhecimento moral. Alguns creem em tudo isto. As duas principais atrações do sentimentalismo moral são, por um lado, dar sentido aos aspectos práticos da moralidade e, por outro lado, encontrar um lugar para a moralidade no interior de uma visão de mundo naturalista. Os desafios correspondentes são explicar as

---

* KAUPPINEN, A. Moral Sentimentalism. In: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2021. Disponível em: https://plato.stanford.edu/archives/win2021/entries/moral-sentimentalism/. Acesso em: 13 jul. 2022.

The following is the translation of the entry on Moral Sentimentalism by Antti Kauppinen, 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/win2021/entries/moral-sentimentalism/. 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 http://plato.stanford.edu/entries/necessary-sufficient/. 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.

aparentes objetividade e normatividade da moralidade. Teorias psicológicas recentes, que enfatizam a centralidade da emoção no pensamento moral, têm promovido um interesse renovado nas éticas sentimentalistas.

1. Os muitos sentimentalismos morais

Considere a seguinte história contada por Frans de Waal:

> J. vive em uma pequena cidade costeira na França, na qual ele é conhecido como o faz-tudo que de fato ele é. Ele pode construir toda uma casa com as suas próprias mãos, pois ele é capacitado para a carpintaria, encanamento, alvenaria, trabalho no telhado, etc. [...] Muito gentil, J. geralmente faz recomendações ou dá uma mão. Um vizinho que ele mal conhecia continuava a perguntar como colocar uma claraboia em seu telhado [...]
> J. passou o dia, até tarde da noite, com seu vizinho, basicamente fazendo o trabalho sozinho (o vizinho mal podia segurar um martelo, ele disse), neste período a esposa do vizinho veio, cozinhou e comeu o almoço com seu marido (a principal refeição na França), sem oferecer nada a J. Ao fim do dia ele obteve sucesso em colocar a claraboia, tendo feito um trabalho especializado que normalmente teria custado mais de seiscentos euros. J. nada pediu, mas quando o mesmo vizinho, alguns dias depois, falou sobre um curso de mergulho e como seria divertido fazê-lo juntos, ele achou que esta seria uma ocasião perfeita para retornar o favor, já que o curso custa em torno de 150 euros. Assim, J. disse que ele adoraria ir, mas infelizmente não tinha o dinheiro em seu orçamento. Agora você já pode adivinhar: o homem foi sozinho (DE WAAL, 2009, p. 174-175).

Se assumirmos que esta história é verdadeira e não deixa de fora nada importante, provavelmente teremos dois tipos de respostas para ela: nós temos um

sentimento negativo de algum tipo com relação ao vizinho e pensamos que o vizinho agiu de maneira errada com relação ao pobre J. Grosso modo, os sentimentalistas pensam que estas duas respostas são intimamente relacionadas, com o sentimento estando no comando. Visto que diferentes teorias na família sentimentalista fazem uso de diferentes respostas, este verbete irá adotar uma definição liberal de um 'sentimento' que inclui atitudes não cognitivas e estados de todos os tipos: emoções, sentimentos, afetos, desejos, planos, e disposições de tê-los.

Esta ideia central levanta muitas questões. Uma é explicativa: por que pensamos que o vizinho fez algo de errado? Os s*entimentalistas explicativos* creem que pensamentos morais são fundamentalmente explicados por sentimentos ou emoções. O segundo tipo de questão é constitutiva: em que consiste o nosso pensamento que o vizinho fez algo de errado? Isto é, que tipo de pensamento é este? É mais como acreditar que Plutão é um planeta ou mais como querer bater em um computador que não coopera? Os *sentimentalistas de juízo*, ao menos em parte, creem que juízos morais são constituídos por respostas emocionais ou não cognitivas ou, alternativamente, são juízos *sobre* respostas emocionais, ou a tendência de algo de dar origem a elas. Alguns sentimentalistas de juízo são também **expressivistas**, sustentando que o significado de termos morais deve ser caracterizado em termos dos estados não cognitivos associados.

Terceiro, assumindo que o entendemos bem, que tipo de fato, se algum, torna verdadeiro o nosso pensamento que o vizinho agiu de maneira errada? É este erro da ação uma projeção dos nossos sentimentos? Se isto é um fato, é como o fato que o quadrado da hipotenusa é igual à soma dos quadrados dos outros dois lados de um triângulo retângulo, ou como o fato que a água é H_2O, ou talvez mais como o fato que comida estragada é repugnante? Para os *sentimentalistas metafísicos*, fatos morais têm a ver com nossas respostas sentimentais, ou *conceitos* morais e avaliativos são conceitos das propriedades de causar ou autorizar respostas sentimentais. Finalmente, nós somos bem confidentes quanto aos nossos juízos. Mas mesmo se assumirmos que sabemos todos os fatos empíricos pertinentes, como sabemos que o que o vizinho fez é errado? E se alguém discordar? Como podemos justificar nosso veredito? Os *sentimentalistas epistêmicos* creem que respostas sentimentais de um tipo apropriado desempenham um papel central na justificação ou compreensão moral.

Estas visões sentimentalistas são logicamente independentes umas das outras. Uma indicação disto é que elas contrastam com diferentes visões. As visões epistêmicas e, possivelmente, explicativas do sentimentalismo contrastam com as

visões **racionalistas** e **intuicionistas**, de acordo com as quais podemos adquirir conhecimentos morais, respectivamente, por raciocínio ou intuição, em vez de através da emoção. Algumas visões do sentimentalismo de juízo, embora não todas, contrastam com explicações cognitivistas de juízos morais. Já as visões sentimentalistas metafísicas contrastam com a teoria do erro e com o realismo moral independente da mente, em suas variantes naturalistas e não naturalistas.

2. Sentimentalismo explicativo

Não é incomum filósofos contemporâneos verem o sentimentalismo como uma teoria do juízo ou fato moral. Mas os sentimentalistas clássicos consideraram como a questão principal a questão da *psicologia moral*: quais características da natureza humana – razão, sentimento ou intuição intelectual – explicam porque aprovamos algo ou culpamos alguém? (*vide Tratado da Natureza Humana* (T) de David Hume, p. 456; *A Teoria dos Sentimentos Morais* (TSM), de Adam Smith, p. 312-313). Alguns de seus principais argumentos foram introduzidos no contexto desta questão explicativa, que se encontra na fronteira entre a filosofia e o que é hoje a ciência empírica da psicologia.

2.1 Teorias do senso moral

Enquanto a importância da emoção no pensamento ético tem sido reconhecida pelos filósofos ao menos desde Aristóteles e Mêncio, as raízes da moderna tradição sentimentalista na ética remete aos debates no início do século dezoito na Grã-Bretanha. Anthony Ashley Cooper, mais conhecido pelo título de Conde de Shaftesbury, introduziu a noção de senso moral [*moral sense*] em seu *Inquiry Concerning Virtue, or Merit*. De acordo com Shaftesbury, o principal objeto da avaliação moral é a afecção [*affection*] ou o motivo por trás da ação. Estas afecções podem, por reflexão, se tornar o objeto de uma afecção de segunda ordem:

> Em uma Criatura capaz de formar Noções gerais das Coisas, não apenas os Seres voltados para fora que se oferecem ao Senso são os Objetos da Afecção; mas as próprias Ações, e as Afecções da Piedade,

> Bondade, Gratidão, e seus Contrários, tendo sido trazidos à Mente pela Reflexão, se tornam Objetos. Assim, por meio deste Senso refletido, surge um novo tipo de Afecção direcionada às próprias Afecções, que já foram sentidas, e agora se tornam o Sujeito de uma nova Inclinação ou Aversão [*Liking or Dislike*] (Shaftesbury 1699-1714, p. 16).

A aprovação moral é explicada por (e talvez consista em) este tipo de inclinação de segunda ordem, o qual Shaftesbury frequentemente designa por "Senso do Certo e Errado". Este senso é inato ou natural e pode ser substituído apenas pelo "hábito e costume contrário" (1699-1714, p. 25), embora ele possa, por vezes, ser superado pela "raiva, luxúria, ou qualquer outra paixão oposta" (1699-1714, p. 35). Na visão de Shaftesbury, o que ele aprova são afecções públicas ou naturais que são os motivos belos e harmoniosos que tendem para o bem de todo o sistema de criaturas racionais. Para ser virtuosa, a criatura deve não apenas fazer as coisas certas, mas exercitar seu senso moral e agir pela própria razão pela qual algo é valoroso e honesto (1699-1714, p. 18). É por isso que mesmo o mais bondoso cachorro não pode ser virtuoso, na medida em que é desprovido da habilidade de refletir quanto às suas afecções e agir no afeto resultante [*resulting affect*].

Frances Hutcheson desenvolveu, até certo ponto, o esboço de Shaftesbury. Seu argumento para a existência de um senso moral recorre a sua rejeição do racionalismo e intuicionismo (*vide* suplemento **Argumentos Antirracionalistas**). Uma vez que não aprovamos ações por causa do autointeresse ou autoevidência, outra coisa deve explicar nossas convicções. De acordo com ele, um senso é uma "determinação da mente, para receber qualquer ideia da presença de um objeto que ocorre em nós, independente de nossa vontade" (1725, p. 90). Há muito mais sensos que aqueles tradicionalmente reconhecidos. Uma vez que ideias morais surgem espontaneamente, Hutcheson conclui que nós temos um senso moral,

> uma determinação das nossas mentes para receber ideias agradáveis ou desagradáveis das ações, quando elas ocorrem para observação, antecedente a quaisquer opiniões de vantagem ou perda para repercutir em nós mesmos a partir delas (1725, p. 100).

O senso moral, mas não as ideias morais, é inato, como pode ser visto nos sentimentos das crianças pequenas quando começam a responder às ações (1725, p. 145-146).

Na primeira visão (e mais conhecida) de Hutcheson, o que o senso moral aprova é acima de tudo "o desejo calmo geral da felicidade dos outros", diferentemente de paixões particulares benevolentes como amor e compaixão (1728, p. 175). Neste quadro, em contraste com Shaftesbury, o motivo para ações moralmente louváveis é a benevolência, não o pensamento que a ação é obrigatória ou que seria aprovada pelo próprio senso moral. Na terminologia de Hutcheson, o senso moral é a fonte de razões justificadoras, mas não de razões estimulantes [*exciting reasons*]. Para um maior desenvolvimento quanto às visões de Hutcheson, confira Bishop (1996).

Algumas das características principais destes primeiros sentimentalistas são compartilhadas pelo modelo contemporâneo do psicólogo Jonathan Haidt (2012). O ponto de partida de Haidt é uma distinção cada vez mais comum na psicologia entre dois tipos de processos cognitivos. Processos de Tipo 1 ou intuitivos são automáticos, sem esforço, rápidos, frequentemente associativos, paralelos, afetivos e frequentemente modulares (isto é, fazendo uso apenas de informações específicas do domínio). Nós somos conscientes apenas de seus resultados, não dos próprios processos (eles são, portanto, paralelos aos "sensos" de Hutcheson). Processos de Tipo 2, "reflexivos" ou "que raciocinam", são, por contraste, processos conscientes, com esforço, lentos, que sobrecarregam a memória [*memory-taxing*], por vezes inferenciais, racionais e lineares. Como é comum na psicologia, Haidt usa o termo "intuição" para crenças resultantes de processos intuitivos.

A afirmação empírica de Haidt é que juízos morais são, na maioria das vezes, intuições aproximadamente causadas por reações instintivas [*gutreactions*], lampejos rápidos e automáticos de afetos. Para embasar esta afirmação, tem-se demonstrado que a manipulação faz diferença no julgamento das pessoas (WHEATLEYAND; HAIDT, 2005), e áreas do cérebro associadas com estas respostas afetivas mostram atividade quando as pessoas fazem julgamentos (*vide* GREENE; HAIDT, 2002; MOLL; OLIVEIRA-SOUZA; ESLINGER, 2003). Ademais, em alguns estudos, as pessoas ficam facilmente estupefatas quando suas visões morais são desafiadas: elas não podem (de acordo com Haidt) fornecer razões para porque desaprovam, por exemplo, incestos inofensivos entre irmãos (HAIDT, 2001). No quadro de certa forma pessimista de Haidt, o papel da razão é, na maioria das vezes, a racionalização legal das reações instintivas pré-existentes, em particular em um contexto social. Da mesma forma, Joshua Greene (2008) alega, baseando-

se em imagem cerebral e dados quanto ao tempo de reação (GREENE *et al*., 2001), bem como em outras coisas, que são em particular os juízos *deontológicos*, tal como o pensamento que é errado matar uma pessoa como um meio para salvar a vida de cinco, que são comandados por reações instintivas afetivas, os quais filósofos como Kant posteriormente tentaram racionalizar. Greene defende que julgamentos deontológicos, em contraste, são o resultado do raciocínio consciente. No entanto, Guy Kahane (2012) argumenta persuasivamente que tão logo olhamos para dados mais abrangentes, nós vemos que a diferença entre processos de Tipo I e processos de Tipo II corresponde à diferença entre juízos intuitivos e contraintuitivos, não à diferença entre os conteúdos deontológicos e utilitários. Enquanto os primeiros sentimentalistas acenavam para a providência divina para explicar o porquê de termos um senso moral e o porquê de aprovarmos coisas como a benevolência, a história de Haidt recorre à evolução. Ele crê que estamos pré-programados para responder, por exemplo, ao sofrimento com a compaixão, à arrogância dos subordinados com desprezo, à trapaça com raiva, e à obscenidade [*impurity*] com repugnância. Estas respostas afetivas culturalmente bem afinadas são, por sua vez, distalmente explicadas por terem sido "adaptações [evolutivas] às ameaças duradoras e oportunidades na vida social" dos nossos ancestrais (HAIDT, 2012).

Um desafio para a visão de Haidt é que, diferentemente das visões sentimentalistas clássicas, ela não distingue entre avaliações morais e não morais. Quando se trata de intuições em seu sentido, ele diz que:

> o cérebro tem um tipo de medidor [*gauge*] (por vezes chamado de "*medidor de afinidade*" [*like-ometer*] que constantemente se move para frente e para trás e estes movimentos, estes movimentos rápidos, influenciam o que quer que venha depois (HAIDT; BJÖRKLUND, 2008, p. 186-187).

Mas nós não consideramos tudo que não gostamos, ou mesmo tudo que nos deixa irritados, como moralmente errado. A história é, pelo menos, incompleta. Ademais, como um psicólogo, Haidt naturalmente não oferece argumentos *a priori* contra a possibilidade de a razão oferecer vereditos morais. Tudo que ele alega é que as pessoas, **geralmente**, não se engajam, antes de qualquer coisa, em raciocinar. Mesmo que isto seja verdade, disto não se segue que os juízos das pessoas não sejam moldados pelo raciocínio ou que elas não sejam sensíveis às razões. De

fato, Daniel Jacobson (2012) sublinha que a evidência que as pessoas carecem de razões para formar seus juízos é, na verdade, muito fraca. No exemplo mais famoso de Haidt, mesmo um incesto inofensivo entre irmãos envolve um **risco** significativo de sofrer repercussões emocionais (RAILTON, 2014, p. 849). O clima no casamento da sua irmã, três anos depois, talvez não seja o mesmo. As pessoas podem não ser capaz de articular imediatamente estas razões, mas isto não mostra que elas não estavam respondendo a estas razões. Em vez de serem meras racionalizações, o raciocínio *post hoc* pode, então, ser um processo de tentar articular as considerações as quais respondem os julgamentos intuitivos de alguém.

Recentemente, Joshua May (2018) tem desafiado a primazia do afeto no pensamento moral a partir de bases empíricas. Ele argumenta que a melhor evidência que nós temos apoia apenas um papel mais modesto para as emoções como facilitando (ou dificultando) o processamento de informação. May nota que manipular emoções incidentais – ou seja, emoções que não são direcionadas ao ou suscitadas pelo objeto de julgamento moral, tais como a repugnância causada por odores desagradáveis (SCHNALL *et al*., 2008) – tem se revelado como tendo efeitos menores. Ele acertadamente aponta que estes efeitos escassos e menores não apoiam quaisquer alegações quanto a um papel dominante ou essencial para o afeto no juízo moral. Seguindo Heidi Maibom (2005), ele também argumenta que problemas no juízo moral dos psicopatas são melhores explicados por déficits racionais, não emocionais.

May sustenta que em vez do mero sentimento, juízos morais são tipicamente explicados pelo raciocínio, embora frequentemente de uma variedade Tipo I automática e inconsciente. Ele apoia esta afirmação ao recorrer a muitos estudos que sugerem que juízos morais são sensíveis às características como intencionalidade (*vide* YOUNG; TSOI, 2013, para uma visão geral) e o uso de força pessoal (*vide* GREENE *et al*., 2009), e, de maneira mais geral, a extensão do envolvimento atribuída ao agente. Como May resume sua visão, "nós frequentemente inferimos rapidamente o estatuto moral de uma ação em parte nos confiando em princípios gerais que identificam como moralmente relevantes várias características dos agentes, ações e resultados [...]. Esta é uma questão de raciocinar por inferência, seja implícita, explícita ou uma combinação das duas ao longo do tempo" (2018, p. 70).

Enquanto os dados e argumentos de May são um corretivo saudável para algumas formas exageradas de sentimentalismo explicativo, tal como o de Haidt, não é claro o quão os sentimentalistas devem se preocupar com eles. Afinal de contas, a principal evidência para o sentimentalismo explicativo não é supostamente

para vir das emoções **incidentais**, mas do fato que nós, digamos, desaprovamos o tipo de coisa que independentemente suscita emoção negativa. O sentimentalista pode de maneira relativamente fácil explicar os alvos de nossas condenações ao recorrer às reações emocionais independentes, enquanto para o racionalista é uma mera coincidência que macacos ficam bravos a respeito do mesmo tipo de coisa que a nossa razão nos diz para reprovar. E é compatível com o sentimentalismo explicativo que as pessoas **primeiro** raciocinam (conscientemente ou não) a respeito de fatos não morais e, depois, respondem com emoções, o que é ainda necessário para atravessar a fenda entre os juízos morais e não morais. Como Hume diz, embora os sentimentos entreguem o veredito final, "a fim de pavimentar o caminho para este sentimento, e prover um discernimento apropriado de seu objeto, achamos que é frequentemente necessário que muitos raciocínios devam preceder, que boas distinções sejam feitas, as justas conclusões retiradas" (1751, p. 15).

2.2 Teorias baseadas na simpatia/empatia

O que é distintivo em David Hume e Adam Smith é que embora eles ocasionalmente usem o termo "senso moral", eles não o tomam como uma capacidade inata e primitiva, mas, antes, pretendem explicar seu funcionamento em termos de mecanismos mais básicos, em particular a partir do que eles chamam de "simpatia". Atualmente, o termo é usado, frequentemente, para um tipo de preocupação com o outro. Mas Hume, Smith e Sophie de Grouchy (1798), de maneiras diferentes, usaram este termo para compartilhar o que o outro sente. O termo contemporâneo para este tipo de mecanismo é **empatia** (para a distinção entre simpatia e empatia, confira Darwall (1998), Sober e Wilson (1998) e Fleischacker (2019)). Não obstante, esta seção irá usar o termo antigo, dado que do que os sentimentalistas clássicos falam não é sempre exatamente "empatia".

Seguindo Hutcheson, Hume rejeita a razão ou o raciocínio como a fonte de distinções morais (juízos ou fatos). Brevemente, ele sustenta que a razão sozinha não pode nos **motivar** da maneira que a moralidade nos motiva (*ceticismo de motivação*) e que apenas o raciocínio não pode nos dizer o que é certo ou errado (*ceticismo de conteúdo*). Para o debate, confira o suplemento **Argumentos Antirracionalistas**.

Na concepção positiva de Hume, a aprovação moral é causada, se não constituída, pelo prazer de um tipo especial, ou uma forma de amor direcionada ao

agente quando seu caráter é considerado imparcialmente:

> É apenas quando um caráter é considerado em geral, sem referência ao nosso interesse particular, que ele causa este sentir ou sentimento [*feeling or sentiment*] e por isso o denominamos moralmente bom ou mau (T, p. 472).

Mas por que a contemplação de um caráter, sem referência ao nosso interesse, nos dá prazer, em primeiro lugar? Não é porque estamos simplesmente equipados com um senso moral inato, mas por causa da operação da simpatia, que oferece uma explicação mais parcimoniosa e sistemática dos fenômenos. Para Hume, seu mecanismo é um tipo de associação analógica dos efeitos de uma paixão para as suas causas, ou vice-versa, que gera a mesma paixão no observador (T, p. 576). Grouchy observou que a importância da memória neste processo tem implicações políticas: o privilegiado que escapou da "escola da dor e adversidade" será deficiente em compaixão (1798, p. 62). Na concepção de Hume, uma vez que os prazeres e as dores dos outros nos são comunicados, nos tornamos de uma maneira especialmente satisfeitos quando contemplamos aqueles traços de caráteres duradouros que fidedignamente satisfaz os outros. Este tipo especial de prazer "denomina" o traço como virtuoso ou vicioso:

> Quando qualquer qualidade, ou caráter, tem a tendência de promover o bem da humanidade, ela nos satisfaz e a aprovamos; pois ela apresenta a ideia viva do prazer; esta ideia nos afeta pela simpatia e é, em si mesma, um tipo de prazer (T, p. 580).

Hume, contudo, nota duas maneiras nas quais nossa aprovação moral diverge do que causa prazer simpático. Primeiro, o que nos vem naturalmente tomarmos o sentimento do próximo e do estimado, aqueles que estão por perto ou são similares a nós e aqueles com os quais nos importamos (esta observação, quanto ao viés inerente à empatia, foi recentemente redescoberta por Prinz (2011) e Bloom (2017)). Todavia, também podemos aprovar as virtudes de estranhos distantes, mesmo as dos inimigos. Segundo, por vezes aprovamos "virtudes em trapos" [*virtue in rags*]: uma pessoa honesta e generosa que, digamos, está presa e, assim, não pode ajudar os outros, não é considerada moralmente inferior que

alguém com os mesmos traços que, de fato, beneficia os outros ao seu redor.

A resposta de Hume para estes desafios é que nós **corrigimos** nossas respostas inicias por viés e da influência da sorte moral, assim como a nossa disposição atual – não porque almejamos um padrão que seja independente de nossas respostas, mas por causa de "várias contradições com os nossos sentimentos na sociedade e em conversação") (T, p. 583),os quais resultam de dependermos de "aparências momentâneas" (T, p. 582). Afinal, quando alguém usa a linguagem moral e chama alguém de "vicioso", "odioso" e "depravado", ele "expressa sentimentos nos quais, ele espera, que toda a audiência irá concordar com ele" (HUME, 1751, p. 75). É esta expectativa de concomitância característica de sentimentos morais que nos impele em direção a um "ponto de vista comum" (ibid.) no julgamento moral, um ponto de vista que abstrai dos nossos interesses particulares, disposição atual e relação com o agente, e foca nos efeitos dos traços característicos da pessoa sobre as pessoas ao seu redor (T, p. 602-603). (*vide* SAYRE-MCCORD, 1994).

Adam Smith apresenta uma teoria diferente quanto à natureza e mecanismos da simpatia e, consequentemente, quanto à aprovação moral e avaliativa. Para Smith, o mecanismo chave da simpatia é se projetar imaginativamente na posição do outro, ou o que agora seria chamado de *simulação* (GORDON, 1995), em vez de um mero contágio, associação ou inferência emocional. Na definição oficial de Smith, a simpatia consiste na concordância entre como alguém considera como o outro de fato se sente e como alguém se sente como resultado de se colocar na posição do outro. A aprovação é o sentimento agradável que resulta de perceber esta coincidência grosseira (TSM, p. 56, nota a), talvez porque ela simultaneamente confirma nossa humanidade comum e a distinção de nossas perspectivas (FLEISCHACKER, 2019, p. 28-31). Embora esta explicação do "senso de propriedade" é por vezes descrita como a teoria do juízo moral de Smith, ele a considera como uma explicação perfeitamente geral da avaliação, abrangendo crença, senso de humor ou decisões de negócios. Quando se trata de juízos de méritos ou deméritos morais, ou as qualidades de merecer elogio ou culpa, o que conta, por um lado, é a nossa simpatia com a **gratidão** e o **ressentimento** daqueles afetados pela ação e, por outro lado, com os motivos do agente. Como Hume, Smith reconhece que a aprovação moral requer o afastamento do ponto de vista pessoal de cada um:

> Mas [o ressentimento e a gratidão], assim como outras paixões da natureza humana, parecem apropriadas e são aprovadas quando o coração de

> todo espectador imparcial inteiramente simpatiza com elas, quando todo espectador indiferente entra inteiramente e as acompanha (TSM, p. 81).

A figura de um espectador imparcial também desempenha um papel chave na explicação de Smith do autodirecionado julgamento moral ou consciência. Porque queremos que os outros simpatizem conosco (e, assim, nos aprovem), nós nos esforçamos para trazer as nossas paixões para um nível no qual um espectador imaginado, imparcial e bem informado, poderia simpatizar (KOPAJTIC, 2020). Antti Kauppinen (2010, p. 236-239; 2014a) argumenta que isto pode ser visto, em termos psicológicos contemporâneos, como uma forma de regulação de emoção por meio da refocalização da atenção e da reavaliação do significado da ação.

A explicação sentimentalista recente de Michael Slote acompanha o fio das teorias baseadas na empatia/simpatia, junto com a ênfase de Hutcheson na benevolência. A alegação de Slote é que a aprovação moral é constituída e explicada por ter empatia com os motivos do agente:

> [S]e as ações do agente refletem uma preocupação empática pelo (bem-estar ou desejos dos) outros, seres empáticos irão se sentir calorosamente ou ternamente para com estes agentes, e este calor ou ternura empaticamente refletem o calor ou ternura empática dos agentes. [...] [E]sta empatia com empatia ... também constitui a aprovação moral, e possivelmente também a admiração, dos agentes e/ou suas ações (SLOTE, 2010, p. 34-35).

Algumas ações das pessoas exibem empatia com os outros. Esta empatia é um sentimento caloroso. Quando temos empatia com o agente, passamos a compartilhar este sentimento caloroso. E este sentimento empático caloroso constitui a aprovação moral. Em contraste, ações não empáticas manifestam uma frieza para com os outros. Uma vez que a aprovação e reprovação moral "entram na constituição dos juízos morais", Slote sustenta que a empatia pode **explicar** as nossas intuições e juízos.

A abordagem de Slote tem sido criticada no interior do próprio campo sentimentalista. Jesse Prinz (2011) observa que esta visão tem dificuldade com relação à reprovação: a falha em ter empatia não é a mesma coisa que a desaprovação,

nem ela têm o mesmo tipo de efeito motivacional. Julia Driver (2011) salienta que pessoas com déficits de empatia podem, não obstante, moralmente aprovar ou reprovar as coisas. A visão de Slote também encara desafios tanto à luz da necessidade como da suficiência, parecendo ser possível que aprovamos ações não empáticas (tal como fazer algo a partir de um sentido de dever, ou fazer a coisa certa pela razão errada), e que há ações empáticas que não aprovamos (tal como ajudar, motivado pela empatia, uma vítimas à custa de muitas outras que estão menos próximas ou são menos parecidas com o agente).

Sentimentos empáticos, contudo, podem entrar na explicação do julgamento moral de maneira mais indireta. A versão do sentimentalismo de Shaun Nichols (2004) compartilha da ambição humeana de explicar o pensamento moral sem assumir uma capacidade ou senso moral inato. Em vez disto, Nichols assume que as pessoas normais têm uma capacidade genérica de ser guiada por normas ou regras de vários tipos, e por certas disposições emocionais não morais, tal como a disposição de ter uma resposta de aversão ao sofrimento dos outros. A hipótese dele é que nós consideramos uma violação de regra como moral (e, assim, errada independentemente do que alguém nos diga como devemos agir), em vez de convencional (algo que deriva seu status proibitivo de alguma autoridade local *de facto*) quando a regra em questão é **apoiada no afeto**, isto é, quando proíbe ações de um tipo para as quais nós temos, de maneira independente, uma reação emocional negativa (NICHOLS, 2004, p. 62).

Uma importante fonte de apoio para esta alegação é que as pessoas que não têm a resposta relevante afetiva, tais como os psicopatas, parecem falhar em distinguir entre regras morais e convencionais, mesmo se suas capacidades racionais forem intactas (*vide* AALTOLA, 2014). Ao mesmo tempo, Nichols sustenta que juízos morais particulares podem ser simples aplicações de regras apoiadas no afeto, sem exigir respostas afetivas *on-line*. Este sentimentalismo explicativo indireto foi antecipado por alguns dos primeiros sentimentalistas. Adam Smith observou que fazemos “a maior parte” dos juízos morais baseando-se em regras gerais, que são para ele baseadas na indução de respostas emocionais para casos particulares (TSM, p. 377). De modo semelhante, Grouchy sustenta que a reflexão gera sentimentos morais gerais e abstratos, os quais obedecemos “sem retomar o caminho pelo qual eles foram primeiramente formados e tudo que os justificou” (1798, p. 112). Edward Westermarck sustentou que “muitos juízos são aplicações de algumas regras gerais aceitas”, cuja aceitação é em última instância explicada pela existência de uma “sanção emocional” na mente do juiz (WESTERMARCK, 1906, p. 6).

Nichols crê que sua explicação dos papéis emocionais também oferece uma explicação para a universalidade de certas normas. Não precisamos assumir que as pessoas têm uma tendência inata para que considerem causar dano aos inocentes como moralmente errado. Em vez disto, é plausível que das várias normas que sociedades diferentes propõem, aquelas que ressoam com as nossas disposições emocionais não morais desfrutam de uma maior "aptidão cultural" (NICHOLS, 2004, p. 127) e, assim, tornam-se mais predominantes ao longo do tempo. Visto que podemos dar uma explicação para (quase todas) regras morais panculturais, assim como para juízos morais sem recorrer a mecanismos inatos dedicados, Nichols (2005) conclui que a moralidade é um subproduto evolutivo, não uma adaptação.

3. Sentimentalismo de Juízo

O sentimentalismo é comumente compreendido com uma tese sobre **julgamentos** ou **conceitos** morais. **Pensamentos** morais envolvem ou se referem aos nossos sentimentos e a **linguagem** moral de uma forma ou de outra expressa ou se refere aos sentimentos. O sentimentalismo de juízo vem tanto em variedades não cognitivistas como cognitivistas.

3.1. Sentimentalismo não cognitivista

Nas visões puramente não cognitivistas, pensamentos morais são constituídos por sentimentos. Pensar que *X* é errado é, ao menos em parte, ter um sentimento negativo direcionado a *X*, ou talvez ter uma atitude positiva de ordem superior direcionada a um sentimento negativo direcionada a *X*. Eis o argumento mais famoso a favor do não cognitivismo:

O Argumento do Internismo
1. Juízos morais são intrinsecamente, ao menos de modo derrotável [*defeasibly*], motivadores (Internismo Fraco).
2. Apenas estados psicológicos não cognitivos, com uma direção de ajuste mundo-para-a-mente são intrinsecamente motivadores; crenças, que possuem

> uma direção de ajuste mente-para-o-mundo, não motivam por si só (Teoria Humeana da Motivação). **3.** Logo, juízos morais são (ao menos em parte) constituídos por estados psicológicos não cognitivos.

O argumento é claramente parecido com os argumentos contra o racionalismo que partem da praticidade (*vide* Suplemento **Antirracionalismo**). A primeira premissa postula uma conexão interna entre o juízo moral e a motivação moral. Os Internistas Fracos acreditam, grosso modo, que julgar que eu moralmente devo parar de fazer *download* de filmes implica que eu tenho ao menos alguma motivação para parar de fazer estes *downloads*. Caso contrário, eu não considero genuinamente a ação como moralmente errada. Minhas convicções morais são manifestas em minhas tendências motivacionais, se não sempre na ação. Por quê? Basicamente porque o ponto de pensar em termos morais não é descobrir alguns fatos sobre o universo, mas nos fazer agir de maneira que nos permitam viver juntos e colher os benefícios da cooperação. Simon Blackburn (1988) e Allan Gibbard (1990) salientam que quaisquer benefícios evolutivos do pensamento moral podem ter dependido de sua praticidade (*vide* Lenman, de 1999, e o verbete sobre **Motivação Moral**).

A segunda premissa do argumento é uma tese em filosofia da mente. Alguns estados mentais, como crenças descritivas, representam como as coisas são e tem uma **direção de ajuste mente-para-o-mundo**. Outros estados mentais não são descritivos, embora eles possam ter o mesmo conteúdo proposicional das crenças. Se você quer que seu carro funcione e nota que ele se recusa a iniciar, seu desejo não tende a deixar de existir, diferentemente de sua crença que seu carro está funcionando (SMITH, 1987). Em vez disso, seu desejo o move a fazer algo que você acredita que *fará* com que o carro funcione, como chamar um mecânico. Ele tem uma **direção de ajuste mundo-para-a-mente**. Embora algumas maneiras particulares de fazer a distinção permaneçam controversas (*vide* SOBEL; COPP, 2001; FROST, 2014), a ideia chave que as próprias crenças descritivas são motivacionalmente inertes é amplamente aceita.

Por fim, a conclusão vem com diferentes forças. Não cognitivistas tradicionais creem que juízos morais são simplesmente estados não cognitivos. Mas que tipo de estados? Eles claramente não são meros desejos, posto que nós podemos desejar coisas que não consideramos como desejáveis. Os primeiros não cognitivistas falavam de um "sentimento ético" especial (AYER, 1936), identificado em termos **fenomenológicos**. Mas não parece haver qualquer caráter fenomenológico comum

a todos os pensamentos morais que possuímos. Assim, em vez disto, não cognitivistas contemporâneos recorrem ao **papel funcional** das atitudes morais, às suas partes distintivas em nossa psicologia moral como um todo. A ideia de Blackburn (1998) é que pensamentos morais envolvem atitudes de ordem superior direcionadas aos desejos e preferências. Quando eu desaprovo moralmente a poluição do meio ambiente, eu simplesmente não desejo não poluir – eu também aplaudo os outros que são avessos a isto e não gosto daqueles que não compartilham meu desejo (BLACKBURN, 2002, p. 125). Quando eu moralmente desaprovo algo, eu também tenho estas atitudes de ordem superior direcionadas às minhas próprias atitudes: eu prefiro não preferir aquilo que desaprovo, o que explica um elo derrotável na motivação.

Gibbard (1990), por sua vez, estritamente identifica juízos morais com juízos sobre a racionalidade da culpa e do ressentimento. Pensar que roubar é errado, por exemplo, é pensar que é racional sentir-se culpado por roubar e ficar ressentido com os outros por roubarem (na ausência de condições como ignorância ou força que escusam o agente). Considerar que algo é racional ou faz sentido, por sua vez, é aceitar normas que permitam isto. Na explicação original de Gibbard, a aceitação de norma é um tipo básico de estado não cognitivo, uma adaptação evolutiva para a coordenação linguisticamente alcançada que não é analisável em termos de outras atitudes (1990, cap. 4). É não cognitiva porque é essencialmente uma tendência motivacional, para agir ou se sentir, de tal modo que estamos preparados para admitir [*avows*] em debates sobre o que fazer. Em um trabalho posterior, ele considera os pensamentos normativos como **planos de contingência** que estabelecem o que fazer em situações atuais e não atuais (GIBBARD, 2003). Estes pensamentos são, de modo derrotável, motivadores, tal como o Argumento do Internismo requer.

3.2 Expressivismo

O argumento acima diz respeito ao juízo moral como um estado psicológico, um tipo de pensamento. Mas o termo "juízo" ["*judgment*"], por vez, também é usado para proferimentos linguísticos. Da mesma forma, o termo "não cognitivismo" ["*non-cognitivism*"] também é usado para uma tese quanto à linguagem moral. A fim de minimizar a confusão, este verbete irá falar de termos morais, proferimentos e sentenças quando se tratar de entidades linguísticas, e reservar o termo "expressivismo" para a tese linguística. Tendo estas estipulações terminológicas em mente, qualquer

argumento para o não cognitivismo pode ser estendido para um argumento a favor do expressivismo, com a adição de algumas premissas enganosamente simples.

> **4.** O significado de uma sentença é determinado pelo estado mental que ela expressa (Psicologismo Metassemântico).
> **5.** Sentenças morais expressam juízos morais.
> **6.** Logo, sentenças morais expressam (ao menos em parte) estados psicológicos não cognitivos.
> **7.** Logo, o significado de sentenças morais é determinado (ao menos em parte) por estados psicológicos não cognitivos.

A premissa 4 é a adição chave. Ela não é a única maneira de chegar a uma teoria amplamente expressivistas na ética – formas anteriores, como o emotivismo (AYER, 1936; STEVENSON, 1944) e o prescritivismo universal (HARE, 1952) dependiam da hipótese problemática que o significado de uma sentença deve ser compreendido com base nos efeitos que ela é usada para alcançar ou no ato de fala que ela é usada para realizar [*perform*]. O Psicologismo Metassemântico não assume esta hipótese, mas simplesmente deriva o significado de expressões linguísticas a partir dos pensamentos que elas expressam, em vez de a partir de suas condições de verdade ou dos mundos possíveis que elas excluem, tal como a semântica padrão o fazem (GIBBARD, 1990, p. 92).

É essencial para compreender o expressivismo que **expressar** um pensamento não é a mesma coisa que **relatar** um pensamento – quando eu digo "O Sol está brilhando", expresso minha crença que o Sol está brilhando, em vez de *relatar* isto, para o qual teria de dizer "Eu acredito que o Sol está brilhando". Mark Schroeder (2010) recentemente tem enfatizado que a tese expressivistas central é que a sentenças "Matar é errado" é relacionada à desaprovação do assassinato da mesma maneira que a sentença "O Sol está brilhando" é relacionada à crença que o Sol está brilhando, o que quer que seja isto. A única diferença é que o estado expresso é não descritivo. Isto, claro, tem consequências semânticas significativas. Visto que estados não descritivos são desprovidos de condições de verdade, expressivistas (puros) notoriamente não podem explicar as relações inferenciais entre estes estados da maneira usual, mas, em vez disto, em termos de acordo e desacordo quanto às atitudes (STEVENSON, 1937; GIBBARD, 2003). Tentativas

de fazer isto tem se mostrado controversas, para dizer o mínimo. Para o debate, confira o verbete sobre **Cognitivismo x Não Cognitivismo Moral**.

Embora seja um traço definitivo do expressivismo que os significados de sentenças morais sejam explicados sem recorrer às suas condições de verdade, é importante notar que expressivistas contemporâneos não negam que faz sentido falar de verdades morais ou mesmo fatos morais. Contudo, para eles falar disto é apenas uma maneira diferente de expressar visões éticas de primeira ordem. Como diz Blackburn, dizer que:

> uma visão ética é verdadeira é apenas reafirmá-la, o mesmo ocorrendo se adicionarmos as palavras pesadas "realmente", "verdade", "fato" e assim por diante. Dizer que é objetivamente verdadeiro é afirmar que sua verdade não varia com o que ocorre de se pensar a respeito e, uma vez mais, esta é uma posição ética interna, de primeira ordem (BLACKBURN, 1998, p. 296).

Fazer com que o caráter guiador da ação da linguagem moral tenha sentido não é, evidentemente, a única motivação para o expressivismo. Outra parte importante do caso a favor do expressivismo é o Argumento Da Superveniência Moral, originalmente proposto por Blackburn (1971, 1985) que, a groso modo, sustenta que os expressivistas podem melhor explicar o porquê que é uma verdade conceitual que mundos descritivamente idênticos devem ser moralmente idênticos. Para mais, ver o verbete sobre Superveniência na Ética.

3.3. Sentimentalismo Cognitivo: Subjetivismo e Disposicionalismo Ideal

Os sentimentalistas não precisam pensar que juízos morais são tão somente sentimentos ou atitudes de algum tipo. Sentimentalistas cognitivos pensam que eles são **crenças** descritivas quanto aos sentimentos ou à disposição de causar sentimentos (irei explorar outras variedades abaixo). A visão semântica relacionada é que sentenças morais são sobre os sentimentos ou disposições para causar sentimentos dos falantes (ou de outra pessoa).

A abordagem semântica mais simples neste sentido é o *Subjetivismo do Falante*, segundo a qual as condições de verdade de um proferimento de "Roubar

é errado" é que o falante desaprova roubar. Esta desaprovação não precisa consistir simplesmente em um sentimento negativo direcionado ao roubo, mas pode também ser pensada como um endosso de um padrão ou norma moral que proíbe o roubo. Em outras palavras, o falante **relata** a mesma atitude que os expressivistas dizem que a sentença **expressa**. Esta visão é sugerida por algumas coisas que Hume diz, em particular as seguintes:

> Quando você declara que qualquer ação ou caráter é vicioso, você não quer dizer nada, além de que partindo da constituição da sua natureza, você tem um afeto ou sentimento [*feeling or sentiment*] de culpa a partir da contemplação disto (T, p. 469).

Para o Subjetivismo do Falante, há um elemento não cognitivo no juízo moral, mas às sentenças podem, não obstante, ser atribuídas uma semântica verocondicional padrão. Assim, sua atração é dar sentido ao internismo de juízo sem a necessidade de semânticas não padrões ou metassemânticas. Uma versão generalizada do Subjetivismo do Falante é o Contextualismo Metaético, de acordo com o qual o padrão relativo ao qual algo é dito como sendo certo ou errado não precisa ser o padrão do próprio falante, mas, em vez disto, é determinado pelo contexto do proferimento, de uma maneira parecida com a qual o contexto fixa o conteúdo de uma sentença tacitamente indexical como "Zheng é alto" (DREIER 1990; BJÖRNSSON; FINLAY, 2010). Desta forma, por exemplo, assim como ao proferir "Zheng é alto", podemos dizer que Zhen é alto **para uma criança**, ao proferir "Roubar é errado", podemos dizer que assim o é para os padrões **que eu e você aceitamos**.

Uma objeção fundamental para estas visões é o Problema do Desacordo Ausente (MOORE, 1912). Suponha que Ann diga que "Comer pessoas é errado" e Beth, uma canibal dedicada, replique "Não, comer pessoas não é errado". De acordo com o Subjetivismo do Falante simples, o que Ann disse é verdadeiro se e somente se Ann desaprova comer pessoas e o que Beth disse é verdadeiro se e somente se Beth não desaprova comer pessoas. É obviamente possível ser simultaneamente verdadeiro que Ann desaprova comer pessoas e que Beth não desaprova, de modo que seus proferimentos não se contradizem. De fato, uma vez que o que Ann diz é verdadeiro neste quadro, parece que pode ser possível para Beth dizer "O que Ann diz é verdadeiro, mas comer pessoas não é errado". Assim, parece que elas não discordam não mais do que se Ann tivesse dito "Meu nome começa com A" e

Beth tivesse dito “Não, meu nome começa com B”. Entretanto, ainda assim elas discordam, de modo que o Subjetivismo do Falante tem um problema.

Subjetivistas e contextualistas contemporâneos são bem cientes deste problema e oferecem diversas sugestões para solucioná-los. Por exemplo, a resposta pragmatista de Gilbert Harman sustenta que quando dizemos que algo é errado, nós **pressupomos** que o ouvinte compartilha de nossos padrões morais (HARMAN, 2000, p. 36). Nossa conversa é sobre o que se segue do nosso padrão, que é algo sobre o qual podemos genuinamente discordar.

O que quer que, de outra forma, façamos da sugestão de Harman, a estratégia pragmática parece indevidamente restringir o escopo do desacordo possível, uma vez que é possível discordar de pessoas que não fazem parte da conversa (BJÖRNSSONAND; FINLAY, 2010). Em resposta, os subjetivistas e os contextualistas, às vezes, abandonam a tese de que o desacordo consiste em crenças contraditórias e explicam o desacordo em termos da noção originalmente não cognitivista de **desacordo na atitude** (*vide* FINLAY, 2017). Assuma, para simplificar, que a desaprovação envolve o desejo de ter o agente punido e que a aprovação envolve o desejo de não ter o agente punido. Então, ainda que os proferimentos de Ann e Beth não se contradigam, elas discordam no sentido que não é possível simultaneamente satisfazer os desejos implicados pela verdade de seus sinceros proferimentos quando elas emitem um veredito quanto ao canibalismo. Se Ann tiver seu desejo satisfeito, Beth não terá e vice-versa.

Outra questão para o Subjetivismo do Falante é que o objeto do pensamento moral não é distintamente avaliativo: se Ellen é presbiteriana e pensa que roubar é errado, o conteúdo do pensamento de Ellen pode ser tão somente que roubar desperta a desaprovação nos presbiterianos. Outra pessoa poderia ter um pensamento com o mesmo conteúdo sem, deste modo, fazer uma avaliação moral do roubo (EGAN, 2012, p. 566). O aspecto avaliativo é **externo** à crença e consiste em algo como se identificar como um Presbiteriano, o que envolve uma atitude, similar ao desejo, no que concerne a evitar aquilo que não se encaixa no código moral presbiteriano. Consequentemente, não há um compromisso com um tipo distinto de fato avaliativo. Na visão de Jamie Dreier (2009), esta é uma vantagem desta explicação, não um problema.

As visões *disposicionalistas ideais* escapam, por meio de dois movimentos, dos problemas com o Subjetivismo do Falante. O primeiro é a idealização: as crenças e as sentenças não se referem às visões do falante, mas às visões de um sujeito apropriadamente **idealizado** (talvez uma visão idealizada do falante). No nível da

forma lógica, alegações de obrigação [*ought claims*] podem ter no local do argumento[53] tácito o padrão, assim como para os contextualistas – mas o padrão não varia com o contexto. O segundo movimento é a disposicionalização: a referência não é aos sentimentos reais das pessoas, mas aos sentimentos que elas **teriam** em circunstâncias apropriadas. Sugestões deste tipo de visão podem ser encontradas em Hume e Smith – recorde que eles falam em **corrigir** os nossos sentimentos tendo como referência o ponto de vista comum ou a resposta do espectador imparcial. A *Teoria do Observador Ideal*, de Roderick Firth (1952), é uma variante mais contemporânea deste tipo de visão.

De uma forma geral, esta visão é a seguinte:

> Julgar que *X* tem uma propriedade moral é crer que qualquer observador ideal teria uma reação eticamente significativa para com *X* sob condições ideais para assim reagir.

Por uma "reação eticamente significativa", Firth (1952, p. 326) quer dizer:

> o tipo de experiência moral que tomamos como evidência, sob condições ideais, para a verdade de nossos juízos éticos.

Ele deixa em aberto se a reação é sentimental, de modo que há possíveis variantes da Teoria do Observador Ideal que não são sentimentalistas. Ao especificar as características do observador ideal, Firth usa uma "metodologia pragmática" de

> examinar os procedimentos que nós realmente consideramos, implicitamente ou explicitamente, como os racionais para decidir questões éticas (*ibid*., p. 332).

---

[53] NT.: O argumento é uma entrada específica na função: o valor que deve ser fornecido para obter o resultado da função. Por exemplo, na função f(x) = x + 1, x é o argumento.

Se realmente desqualificamos os vereditos éticos de alguém por ignorância de fatos não morais, falha de vivamente imaginar como algo realmente seria, parcialidade, emoções não morais e inconsistência, isto mostra que nós implicitamente consideramos juízos morais como válidos apenas quando feitos por um julgador onisciente, que tudo percebe, imparcial, desapaixonado, consistente, mas, por outro lado, um julgador normal (*ibid*., p. 333-345). Estas são, portanto, as características do observador ideal.

As visões *disposicionalistas ideais* evitam o Problema do Desacordo Ausente. Por exemplo, se Ann acredita que qualquer espectador informado imparcial iria desaprovar *X* e Beth acredita que não é o caso que qualquer espectador informado imparcial não iria desaprovar *X*, suas crenças se contradizem e elas discordam frontalmente.

Contudo, esta vitória para as visões idealizadas é alcançada pelo deslocamento das atitudes reais das próprias pessoas, o que leva ao Problema da Motivação Ausente. A atração que o subjetivismo simples, contextualismo e não cognitivismo compartilham é que se os pensamentos morais consistem em atitudes ou crenças quanto aos padrões que se endossa, a suposta conexão não contingente deles com a motivação não é problemática. Mas como crenças sobre disposições, que causam respostas sentimentais em sujeitos ideais, podem necessariamente motivar pensadores comuns, dada a Teoria Humeana da Motivação (MACKIE, 1980, p. 69)? Alguns podem ter o desejo de fazer o que um espectador imparcial aprova, mas o que dizer daqueles que não tem este desejo? Como Richard Joyce (2013) observou, parece que alguém pode estar tão desmotivado por estes pensamentos como pela crença que vikings bêbados zombariam da realização de uma ação, e isto seria plenamente racional.

Michael Smith (1994, 1997) oferece uma resposta distintamente **racionalista** a este desafio, ao recorrer à ideia de que uma agente é mais racional, em termos de coerência, se ela é motivada para fazer o que ela acredita que seu conselheiro, informado e idealmente racional, gostaria que ela fizesse. Se uma visão sentimentalista poderia fazer um movimento similar é algo que merece uma exploração posterior. Por exemplo, uma sugestão inspirada em Adam Smith seria que uma agente que crê que qualquer espectador imparcial iria desejar que o agente fizesse *A* e este agente falhasse em desejar A é menos coerente que uma agente que combina a crença com o desejo que *A*. Isto assume que as nossas crenças quanto às atitudes de um espectador ideal são inerentemente impositivas [*authoritative*] para nós (diferentemente de crenças quanto aos Vikings), como de fato Smith pensava.

3.4 Sentimentalismo cognitivista: a teoria da sensibilidade e o neossentimentalismo

As visões cognitivistas sentimentalistas, debatidas na seção anterior, apelaram para crenças quanto às disposições de causar respostas sentimentais em certos tipos de sujeitos. O outro tipo principal de visão cognitivista, a teoria da sensibilidade, considera juízos como crenças sobre respostas merecidas (WIGGINS, 1987; MCDOWELL, 1998). A teoria da sensibilidade começa com uma analogia com conceitos de qualidade secundária, tais como os conceitos de cor. Estes conceitos, de acordo com muitos, são conceitos de qualidades dependentes da mente. Os teóricos da sensibilidade acreditam que eles são parcialmente análogos com conceitos de valores, a diferença sendo que o valor é

> concebido como sendo não meramente para suscitar a "atitude" apropriada [...] mas para **merecê-la** (MCDOWELL, 1985, p. 143).

Quando concebemos o valor desta maneira, nós percebemos que há uma razão para, digamos, admirar ou emular alguém (MCDOWELL, 1979).

Este estado de perceber ou conceber que há uma razão é uma crença de um tipo único e controverso. De acordo com John McDowell (1978), ela pode explicar a ação virtuosa de uma pessoa sem a ajuda de um desejo relacionado exercendo um papel causal (embora possamos devidamente atribuir um desejo a uma pessoa como uma consequência de ela estar motivada pela percepção de uma razão). Assim, McDowell rejeita a Teoria Humeana da Motivação (premissa 2 do Argumento do Internismo). Seus argumentos são inspirados por sua leitura de Aristóteles, em vez da tradição sentimentalista da primeira filosofia moderna. Grosso modo, uma pessoa virtuosa vê as situações de uma maneira distintiva (MCDOWELL, 1979, p. 73), de modo que certas características, como a necessidade de alguém de saber algo, são **salientes** para ela, enquanto outras preocupações são **silenciadas** e, portanto, não se destacam como pedindo ação. Esta "visão moral" (MURDOCH, 1970), então, é suficiente para explicar o que o agente faz.

Uma objeção familiar é que é possível para alguém compartilhar a crença da pessoa virtuosa, ou a "concepção da situação", mas, não obstante, não ter a motivação. McDowell (1978, p. 87) nega isto:

> [A]s concepções relevantes não são possuídas exceto por aqueles cujas vontades são apropriadamente influenciadas.

Um argumento a favor disto é que há uma diferença entre ser virtuoso e ser casto [*continent*]. Ambos os tipos dos agentes fazem a coisa certa – por exemplo, ambos permanecem fiéis aos seus parceiros. Mas o agente casto tem de lutar com desejos concorrentes que obscurecem sua atenção quanto ao bom e no nobre [*noble and the fine*]. Ele tem de reunir força de vontade para evitar a tentação. A diferença entre a pessoa casta e a virtuosa, neste quadro, não é na primeira instância uma diferença conotativa, mas uma diferença quanto ao que é saliente, o que é uma diferença simultaneamente cognitiva e afetiva. Nas palavras de David McNaughton, "uma maneira de ver uma situação pode ela mesma ser uma maneira de sentir e cuidar" (MCNAUGHTON, 1988, p. 113). Uma consequência desta visão é que o processo de vir a apreciar razões práticas irá envolver a modelagem das sensibilidades motivacionais do agente, podendo ser aparentada a uma conversão não racional, em vez de a uma deliberação racional a partir de motivos existentes (MCDOWELL, 1995, p. 100-101).

Teorias da sensibilidade pertencem à categoria que Justin D'Arms e Daniel Jacobson (2000a) rotularam de *neossentimentalismo*. Oficialmente, esta é a visão que *conceitos* avaliativos são conceitos da adequação de respostas sentimentais (como será debatido na seção 4.3, os neossentimentalistas frequentemente fazem alegações metafísicas relacionadas quanto às propriedades avaliativas). Isto é, pensar que algo é vergonhoso consiste em pensar que a vergonha é apropriada, ou que há razão suficiente para isto. Uma motivação chave para o neossentimentalismo é que fazer um julgamento quanto ao desejo ou vergonha [*desirability or shamefulness*] é diferente de desejar ou estar envergonhado. O sentido de usar e introduzir conceitos normativos é **guiar** as nossas atitudes, em vez de apenas expressá-las ou reportá-las (*vide* verbete sobre **Teorias das Atitudes Apropriadas do Valor**).

3.5. Abordagens Híbridas e Pluralistas

Teorias híbridas solucionam o problema de como se encaixar os aspectos práticos e representacionais dos juízos morais argumentando que tanto estados cognitivos como não cognitivos estão de alguma forma em jogo: juízos morais têm

aspectos tanto sentimentais como não sentimentais. Diversas opções têm sido exploradas recentemente, incluindo o expressivismo híbrido, teorias dos estados híbridos e o pluralismo do pensamento moral, que serão debatidas abaixo.

O *expressivismo híbrido* vem em muitas variedades. Talvez a mais conhecida seja o "expressivismo ecumênico" de Michael Ridge (2006). Considere o juízo que eu devo dizer a verdade para o juiz. Como primeira aproximação, este juízo consiste em aprovar as ações de um certo tipo e a crença que dizer a verdade ao juiz é uma ação de um certo tipo (*vide* BARKER, 2000). Talvez eu aprove ações que maximizem a utilidade e creia que dizer a verdade ao juiz maximize a utilidade. Para Ridge (2014), o elemento não cognitivo abrange uma perspectiva normativa que admite um conjunto de padrões como impositivos para o raciocínio prático (neste caso, um raciocínio utilitarista) e o elemento cognitivo abrange uma crença comum que estes padrões exigem que se diga a verdade.

Ridge argumenta que este tipo de visão explica as atrações do não cognitivismo, principalmente ao evitar compromissos ontológicos problemáticos com fatos morais não deflacionários e ao explicar a praticidade. Quando eles tem uma perspectiva normativa que envolve uma atitude semelhante ao desejo para com o que eles acreditam como requerendo certos padrões, os argentes racionais práticos formam um desejo instrumental de realizar estas ações. O elemento não cognitivo do juízo, por sua vez, serve para explicar a composicionalidade e as características inferenciais da linguagem moral no estilo verocondicional familiar. Para outras versões do expressivismo que envolvem expressar estados semelhantes ao desejo e crenças apropriadamente relacionadas, ver especialmente Boisvert (2008), Toppinen (2013), Schroeder (2013) e Fletcherand e Ridge (2014).

A explicação de Jesse Prinz (2007) de juízos morais pode ser classificada como uma *teoria do estado híbrido*. O cerne de sua visão é a sua teoria das emoções (PRINZ, 2004), de acordo com a qual as emoções são percepções de padrões de mudanças corporais que representam aquilo em resposta ao qual elas são "configuradas para serem iniciadas" – o que é a função delas detectar – da mesma maneira como quaisquer outros estados psicológicos o fazem nas teorias teleológicas da representação. Para Prinz, um **sentimento** moral é uma disposição de responder a certas ações com uma série de censura ou louvor, direcionada aos outros ou a nós mesmos, constituinte de emoções, tais como culpa, desprezo, raiva e gratidão. Um julgamento moral consiste na emoção que resulta da ativação de um sentimento, tal como a raiva pelo ato de roubar ou a vergonha de fugir. A raiva representa o roubo como sendo algo que causa desaprovação no julgador – isto é, como

moralmente errado, de acordo com a metafísica moral relativista de Prinz (*vide* sec. 4.1). Isto significa que o juízo pode ser verdadeiro ou falso (relativo ao julgador). Isto também motiva a punição do agente. Outras emoções e, assim, juízos, têm efeitos motivacionais diferentes – a repugnância pode motivar a evitar o agente. Um desafio evidente para esta visão é, assim, a aparente possibilidade de fazermos juízos morais não emocionais, assim como a de ter emoções morais não julgadoras.

Por fim, o pluralismo de pensamento moral, segundo o qual pensamentos morais diferentes são constituídos por tipos diferentes de estados psicológicos, é uma abordagem muito recente, embora Elizabeth Radcliffe (2006) argumente que Hume já distinguia dois tipos de pensamento moral. Linda Zagzebski (2003) e Uriah Kriegel (2012) argumentam de diferentes maneiras que há dois tipos de juízos morais, um deles contento um elemento afetivo. A sugestão de Kriegel é que juízos morais resultantes de processos de Tipo I, automáticos e não conscientes, consistem naquilo que Tamar Gendler (2008) rotulou de "*aliefs*", estados cognitivos que afetam o nosso comportamento, ainda que eles vão de encontro aos nossos juízos conscientes. Juízos morais que resultam do pensamento Tipo II, consciente e esforçado, são para Kriegel crenças ordinárias com uma fenomenologia objetivista. Dado que Kriegel aceita que crenças ordinárias por si só não motivam, na sua visão o internismo é verdadeiro apenas quanto aos *aliefs* morais. Acerca do ceticismo quanto aos *aliefs*, confira o trabalho de Sinhababu (2017, p. 129-134).

De maneira similar, Antti Kauppinen (2013, 2015) distingue entre aparências ou **intuições** morais, que são constituídas por manifestações (tipicamente emocionais) de sentimentos que esperamos que os outros compartilhem, e **juízos** morais, que são (implicitamente) crenças quanto ao que seria permitido, requerido ou recomendado pelos padrões que um sujeito ideal endossaria. Assim como com outras aparências, a existência de pensamentos morais que não são juízos é mais clara nos casos nos quais eles entram em conflito com o juízo, tal como quando aparece para Huck Finn que é errado revelar onde Jim está, embora ele acredite que ele deve fazer isto. Kauppinen mantém que aparências [*seemings*] morais tem um caráter fenomenológico e motivacional distinto e diverso (em contraste com, digamos, intuições matemáticas), o que é melhor explicado por eles serem, por natureza, sentimentais. Portanto, o internismo motivacional é verdadeiro a respeito de aparências morais, tal como Kriegel sustenta para *aliefs* morais.

Um contraste importante com Krigel é que para Kauppinen, aparências morais não são elas mesmas juízos, mas explicam e potencialmente justificam crenças morais. Aparências morais são, portanto, tomadas como paralelas às

aparências perceptuais – estados não doxásticos que atraem consentimento aos seus conteúdos proposicionais (*vide* sec. 5.1). Graham Oddie (2005) defende a visão segundo a qual **desejos** constituem aparências de valor, desempenhando um papel epistêmico similar. Visto que aparências frequentemente causam crenças morais, os dois tipos de pensamentos frequentemente co-ocorrem. Mas crenças morais podem ser formadas e sustentadas independentemente de aparências morais. Isto explica a força das objeções externistas motivacionais e a possibilidade do amoralismo (embora, na visão de Kauppinen, agentes morais completamente competentes são motivacionalmente e deliberadamente sensíveis às crenças com estes conteúdos). Visto que as crenças expressas por juízos morais são crenças descritivas comuns, não há necessidade para uma semântica moral separada. As visões pluralistas, portanto, fazem a separação entre o sentimentalismo cognitivista e não cognitivista: cada uma é, em termos gerais, verdadeira quanto a um tipo importante de pensamento moral (em contraste com as visões híbridas, segundo as quais **cada** pensamento tem tanto elementos cognitivistas como não cognitivistas).

4. Metafísica Moral

É um pensamento atraente que fatos e propriedades morais (e outros avaliativos) não estão brutalmente aí fora independentemente do pensamento e sensibilidade humana. Tome a propriedade avaliativa de ser engraçado. Como algo poderia ser divertido, se nenhum ser humano nunca tivesse sido divertido por ele? Como algo poderia ser ultrajante se falhasse em gerar qualquer ultraje quando fosse conhecido, mesmo naqueles que se importam com a vítima? Para os sentimentalistas, o valor, incluindo o valor moral, é **antropocêntrico** (D'ARMS; JACOBSON, 2006).

Os sentimentalistas concordam com os teóricos do erro que fatos morais *sui generis*, não naturais, seriam excêntricos e que fatos naturais independentes da mente são inapropriados para o papel de fatos morais. Sentimentalistas de juízo não cognitivos tratam fatos morais como **projeções** de atitudes morais. Embora haja uma razão excelente para pensar que Hume não era um projetivista, ele fez uma das formulações clássicas ao dizer que o gosto

> tem uma faculdade produtiva, adornando e manchando todos os objetos naturais, emprestada do sentimento

> interno, que ergue de certa maneira uma nova criação (1751, p. 88).

Tomamos uma resposta sentimental em nós mesmo como, digamos, a inveja e atribuímos ao objeto mundano uma característica que ele realmente não possui, tal como a de ser invejável.

Os teóricos do erro tratam esta projeção da mesma forma que a atribuição de agência aos planetas por parte de algumas culturas: uma crença falsa predizível. Expressivistas contemporâneos, por contraste, creem que a atribuição de fatos e propriedades morais tem um propósito. Ela nos permite expressar nossos comprometimentos de uma maneira que fornece um ponto focal para o desacordo e o debate. Como debatido acima, nas visões deflacionárias da verdade e dos fatos, trata-se de uma questão de primeira ordem se algumas atribuições incorrem em erro ou não (BLACKBURN, 1998).

O mais interessante para os metafísicos morais é que muitos sentimentalistas de juízo cognitivistas pensam que juízos morais se referem a fatos morais e avaliativos que são metafisicamente determinados por respostas sentimentais de algum tipo. De acordo com estas visões, propriedades morais são, de uma maneira ou de outra, dependentes de resposta. Para mais, ver o Suplemento Dependência de Respostas.

4.1. Subjetivismo e Relativismo

A maneira mais simples de ligar propriedade morais às nossas respostas emocionais ou conativas seria dizer que o que torna algo errado é que ele é desaprovado e algo ruim é que não é desejado. Ninguém sustenta uma visão tão elementar, pois isto implica que, por exemplo, um assassino não detectado não é moralmente errado. Isto também deixa indeterminado qual desaprovação está em debate – e se uma pessoa desaprova uma ação, mas outra aprova? O que irei chamar de Subjetivismo Simples Disposicionalista avança neste ponto ao indexar propriedades morais a sujeitos e ao se referir às **disposições** para causar respostas, não às respostas reais. O que torna algo errado-para-Mary, por exemplo, é que Mary iria desaprová-la, se estivesse ciente disto. Uma proposta relativista relacionada é que algo é errado-para-Mary se e somente se Mary aceita um padrão ou esquema normativo que proíbe isto (HARMAN, 1975).

De acordo com Jesse Prinz (2007, p. 92)

> Uma ação tem a propriedade de ser moralmente errada (certa) apensa se houver um observador que tem um sentimento de desaprovação (aprovação) direcionado a ela.

Recorde que, para Prinz, um sentimento é uma **disposição** para sentir várias emoções, de modo que o erro não depende de alguém realmente ter uma reação negativa para com algo. Mas esta alegação é altamente implausível, dado que o erro não é indexado. Isto é verdade quanto a quase tudo que alguém irá desaprovar e outra pessoa irá aprovar. Logo, a mutilação genital feminina, por exemplo, é simultaneamente (em termos absolutos) **tanto** certa como errada. Este é um custo teórico exorbitante.

O Subjetivismo Simples Disposicionalista parece implicar que se começarmos a aprovar a escravidão, a escravidão viria a ser moralmente correta – seria verdade que iríamos, digamos, responder positivamente quando a encontrássemos. Mas, certamente, a descrição correta deste cenário não seria que a escravidão se tornou correta, mas que nos tornamos pessoas piores (BROAD, 1944/5, p. 151). Eu chamo este de o *Problema da Rigidez Ausente*. O *Problema da Falibilidade Ausente*, por sua vez, é que uma visão subjetivista simples parece deixar espaço insuficiente para *erros* quanto ao valor. Se o que é correto é determinado por aquilo que aprovo, eu posso estar errado quanto ao que é correto apenas se estiver errado quanto ao que de fato aprovo. Mas é muito implausível que o remédio para erros morais seja uma melhor introspecção.

4.2. Disposicionalismo Ideal

Em paralelo ao caso de juízos morais, os problemas com o Subjetivismo Simples motivam o movimento para o Disposicionalismo Ideal. Segundo este tipo de perspectiva, a extensão das propriedades morais ou avaliativas é determinada pelas respostas sentimentais de sujeitos idealizados de algum tipo, ou pelas respostas sob condições ideais. Consequentemente, o Problema da Falibilidade Ausente é facilmente contornável: sem dúvida, podemos estar errados quanto ao que observador ideal iria aprovar.

As visões idealistas disposicionais também evitam o Problema da Rigidez

Ausente. A depender dos detalhes, pode ser suficiente idealizar o sujeito dos sentimentos: talvez nenhum espectador imparcial aprovaria a escravidão, ainda que a maioria das pessoas tivessem uma sensibilidade pró-escravidão. Em caso de o processo de idealização ser dependente da trajetória (de modo que o resultado dependa dos sentimentos iniciais não idealizados do sujeito), disposicionalistas ideais poderiam fazer um movimento de rigidez (WIGGINS 1987, p. 206). Eles podem dizer que o ponto de partida da idealização é constituído pelos sentimentos normais **reais**, o que inclui um forte desejo de não estar à mercê da boa vontade dos outros, entre outras coisas. Assim, observadores ideais rigidificados desaprovariam a escravidão, ainda que tenha se tornado estatisticamente normal aprová-la. Peter Railton (2003) objeta à rigidificação, porque para ele o que é intrinsecamente bom ou mau para alguém é **relacional**, dependendo da combinação entre suas respostas e seus objetos, assim como ocorre com a doçura e o amargor. A mudança na sensibilidade das pessoas **iria** modificar o que é bom (ou doce) para elas. Ainda assim, algo como a crueldade não iria se tornar correto mesmo que sua aprovação fosse generalizada, uma vez que, para Railton, a sua incorreção depende da maldade das suas consequências, não da nossa desaprovação, e a dor à qual as pessoas são avessas seria, ainda assim, má para elas.

O Idealismo Disposicional vem em muitas variedades que frequentemente correspondem às visões quanto aos conteúdos dos juízos morais (FIRTH, 1952; LEWIS, 1989; SMITH, 1994). Ainda que eles evitem os problemas da falibilidade e rigidez, há uma série de outros desafios. Cinco serão aqui debatidos.

Primeiro, o Dilema de Eutífron deriva seu nome do diálogo platônico no qual Sócrates argumenta contra uma versão da teoria do Comando Divino. Aplicado ao subjetivismo, o dilema pode ser assim apresentado:

> **1.** Uma teoria subjetivista define o sujeito cujas respostas determinam a extensão das propriedades avaliativas, sejam em termos cuja aplicação pressupõe verdades avaliativas dependentes da mente (abreviadamente, termos avaliativos) ou em termos não avaliativos.
> **2.** Se o sujeito é definido em termos avaliativos, a teoria colapsa no realismo independente da mente: ela pressupõe a existência de ao menos algumas verdades avaliativas independentes da mente.

> **3.** Se o sujeito é definido em termos não avaliativo, sua resposta será moralmente arbitrária e desprovida de autoridade para determinar a extensão de propriedades avaliativas.
> **4.** Logo, uma teoria subjetivista ou é autodestrutiva (pois colapsa no realismo) ou falsa (pois produz extensões arbitrárias para as propriedades morais).

É claro que concepções subjetivistas com qualquer tipo de ambição redutiva não pode abraçar o primeiro chifre do dilema (ver abaixo as explicações não reducionistas). Logo, é o segundo chifre do dilema que elas devem encarar. Qual é exatamente o problema com ele? Segundo RussShafer-Landau (2003, p. 41), é que

> pode ser impossível elaborar o conjunto de restrições [não avaliativas] quanto à formação de atitudes tal que as atitudes emergentes geram prescrições que correspondam às nossas visões quanto ao que paradigmaticamente constitui o comportamento moral e imoral.

O desafio é que os observadores ideais, quando definidos em termos não avaliativos, podem aprovar coisas más, como limpeza étnica, o que iria contraintuitivamente implicar que a limpeza étnica seria moralmente correta. Contudo, Shafer-Landau admite que não há um argumento definitivo para mostrar que qualquer idealização em termos não avaliativos irá gerar resultados contraintuitivos.

Talvez o desafio real que o Eutífron estabelece diz respeito à *autoridade normativa* das respostas de qualquer tipo de sujeito definido em termos não avaliativos. Uma maneira de desenvolver este desafio pode ser chamado de *o Problema da Normatividade Ausente*. Muitos acreditam que a correção e incorreção moral são características *categorialmente fornecedoras de razão* das ações. Todos, independente de seus interesses e desejos, têm uma razão para abster-se de comportamentos moralmente errados. Suponha, então, que estar errado consiste em estar desaprovado por qualquer observador ideal. Em uma visão reducionista, esta é uma propriedade natural da ação, especificada em termos não normativos. Eis o desafio: por que seria esta propriedade natural categorialmente fornecedora de razão? Por que as pessoas deveriam se importar em apenas fazer aquilo que não seria desaprovado por um espectador hipotético?

Como Mark Johnston diz,

> Dizer que algo seria valorizado sob condições *K* não é, deste modo, recomendá-lo, mas apenas fazer uma observação descritiva quanto a sua relação com certas condições psicológicas (JOHNSTON, 1989, p. 157; *vide* BLACKBURN, 1993, p. 274).

Estes desafios podem ser maneiras de entender a força do famoso Argumento da Questão Aberta de G. E. Moore (1903): dado que faz sentido conceder que algum *X* seria desaprovado por qualquer observador imparcial, informado, mas de resto normal, mas, no entanto, negar que *X* é moralmente errado, há uma razão para pensar que ser desaprovado por um observador ideal não é a mesma coisa que ser moralmente errado. Certamente o Argumento da Questão Aberta é controverso e há uma série de respostas padrões: talvez a identidade das propriedades seja *a posteriori,* como aquela entre água e H_2O (BRINK, 1989), ou talvez seja uma verdade conceitual não óbvia que nós podemos identificar as platitudes circundantes aos nossos conceitos (SMITH, 1994). Confira o verbete sobre **naturalismo moral**.

A terceira questão é o problema da indeterminação. Dada uma caracterização do observador ideal, há de fato uma ação que seria aprovada por **qualquer** observador ideal, dada uma divergência de atitudes pré-idealizadas (BRANDT, 1955, p. 408)? Se observadores ideais discordam, não há, de acordo com a teoria, uma questão de fato. Este problema deriva de uma característica essencial das teorias disposicionalistas idealizadas: a idealização começa com como as pessoas realmente são e o processo pode não garantir convergência nos aspectos relevantes.

Finalmente, o Problema da Não Motivação Idealizada é que não é óbvio que o movimento a partir das respostas ou disposições reais para as disposições ideais tem um conjunto de razões que seja consistente com as principais ideias da teoria da dependência de resposta. David Enoch (2005) recentemente propôs um caso convincente para este problema. Ele nota que se há fatos independentes de respostas, faz sentido privilegiar respostas de certos tipos de sujeitos em certas condições, porque estas características dos sujeitos e das condições são propícias a rastrear os fatos independentes das respostas. Mas obviamente esta não é uma opção disponível para o disposicionalistas ideal.

Enoch alega que não há alternativa racional real que não seja ad hoc ou que não dependa tacitamente do realismo independente da resposta. Por exemplo,

recorrer a nossa prática real de justificação, na qual nós podemos privilegiar respostas baseadas, por exemplo, no conhecimento imaginativo completo, não é uma opção, visto que esta prática é ela mesma melhor explicada pela suposição subjacente que há fatos morais independentes da mente (ENOCH, 2005, p. 774). Contudo, Hume pode objetar que a base racional para recuar das nossas repostas sentimentais reais é evitar "as contradições contínuas na sociedade e na conversação" e outros problemas práticos (*vide* seç. 2.2). Se isto é correto, nossa prática real de justificação e o lugar privilegiado que ela fornece às respostas reflexivamente corrigidas possuem, então, uma explicação e base racional que não pressupõem fatos avaliativos independentes da mente. Neste sentido, Kauppinen (2014b, p. 580) sustenta que "uma perspectiva avaliativa é ideal ou ótima quando adotá-la de maneira mais confiável evita os problemas práticos caudados pelas atitudes não corrigidas" (*vide* SOBEL, 2009).

4.3. A Teoria da Sensibilidade e o Neossentimentalismo

Para as teorias da sensibilidade, uma propriedade avaliativa não é uma disposição para **causar** respostas em nós, mas a propriedade de **merecer** uma resposta nossa. Uma versão, baseada na teoria da sensibilidade, da equação da dependência de respostas parece com isso:

> *X* é *M* se e somente se **X merece R**/ qualquer sujeito **virtuoso S** responderia a **X** com **R** em circunstâncias **ideais**.

Trata-se de uma reclamação óbvia e familiar contra as visões deste tipo que a análise *não é informativa*. Afinal de contas, termos normativos ocorrem em ambos os lados da equação: o lado direito se refere aos sujeitos *meritórios* e *virtuosos*. Não há tentativa de reduzir propriedades avaliativas a propriedades não avaliativas. Nenhuma das respostas nem das propriedades têm prioridade metafísica, mas são "irmãs" (MCDOWELL, 1987). Um dos argumentos a favor disto é a alegação que a resposta relevante não pode ser identificada independentemente do (conceito da) propriedade (WIGGINS, 1987, p. 195). Não obstante, as teorias da sensibilidade defendem que a elucidação não redutiva oferece uma vantagem frente às visões da qualidade primária, segundo as quais propriedades morais não têm qualquer

relação essencial com nós mesmos (*vide* CUNEO, 2001).

As teorias da sensibilidade são uma espécie do que frequentemente se chama de Análise das Atitudes Apropriadas [*Fitting Attiudes*] (AA) do Valor (RABINOWICZ; RØNNOW-RASMUSSEN, 2004). A forma geral de uma análise AA é a seguinte:

> X é E se e somente se, e porque, R é uma resposta **adequada** ou **apropriada** para **X /** há uma **razão suficiente** para responder a **X** com **R**.

O "porquê" indica que a direção da explicação metafísica é da adequação da atitude para a propriedade, não o contrário: algo é bom porque é apropriado desejá-lo, não apropriado desejá-lo porque é bom. Isto exclui visões nas quais o valor e as atitudes apropriadas caminham juntas, mas defendem que é o valor que torna a atitude apropriada, não vice-versa. Alguns identificam as análises AA com o neossentimentalismo, mas esta é uma classificação não proveitosa, pois muitas formas das análises AA consideração a adequação ou apropriação das atitudes em si mesmas ou como um fato normativo independente da mente que pode ser intelectualmente intuído (*vide* EWING, 1948) ou como uma verdade determinada pelo que seria racionalmente ou razoavelmente aceito (*vide* SCANLON, 1998). Estas visões AA, portanto, se afastam significativamente das principais motivações para o sentimentalismo e são melhores classificadas como variações AA do não naturalismo e construtivismo.

As características específicas da análise AA sentimentalista revelam-se mais notavelmente nas respostas que os sentimentalistas podem dar para o eco metafísico do Problema da Confluência, a saber, o problema dos Tipos Errados de Razões. Este é o problema quanto ao que torna uma resposta apropriada ou dá razão suficiente para ela pode ser algo que nada tem a ver com o valor relevante. Se um demônio do mal ameaça te bater a menos que você o admire por isso, você pode ter uma razão suficiente para admirar o demônio (RABINOWICZ; RØNNOW-RASMUSSEN, 2004). Mas isto não significa que o demônio é admirável – pelo contrário. No jargão, a admiração não é uma resposta **apropriada** para o demônio, visto que é adequada pelo tipo errado de razão.

Há várias razões para distinguir entre tipos certos e errados de razões para atitudes, sendo que nem todas são sentimentalistas (*vide* PARFIT, 2011). Justin D'Arms e Daniel Jacobson (2000b) apresentam uma alternativa distintamente

sentimentalista. A solução deles recorre aos padrões *aléticos* de correção derivados do conteúdo das atitudes em questão. Assumindo que uma emoção tem um componente cognitivo que apresenta ou constrói as coisas como sendo de uma certa maneira (*vide* seç. 5.1), ela pode ser acurada ou inacurada, a depender se as coisas são desta maneira. O tipo correto de razões contra ou a favor da atitude são aquelas que dependem de sua acurácia. Como eles dizem,

> Um episódio emocional apresenta o seu objeto como tendo certas características avaliativas; ele é inapropriado [...] quando o seu objeto é desprovido dessas características (2000b, p. 73).

O exemplo que D'Arms e Jacobson usam é a inveja. De acordo com eles, ela (em geral) apresenta um rival sob uma perspectiva negativa em virtude de ter uma posse desejável. O tipo certo de razões, a depender se um rival é invejável, são aquelas que tem a ver com o que ele possui e o grau de desejo relativo ao que se possui. Se um colega ganha uma promoção que você queria, a inveja pode ser apropriada neste sentido e, consequentemente, dada uma abordagem baseada em atitudes apropriadas, o colega será invejável. Este é o caso ainda que a inveja seja moralmente ou prudencialmente inadequada – este último tipo de razões não depende da adequação da inveja.

Um desenvolvimento alternativo do mesmo tipo de ideia pode ser encontrado no trabalho recente de Christine Tappolet (2016). Ela defende que em vez de formular a adequação em termos de razões ou requisitos normativos de qualquer tipo, os neossentimentalistas podem sustentar simplesmente que, digamos, Pierre é admirável se e somente se a admiração é *correta* em resposta a Pierre quando o contemplamos (2016, p. 87). Aqui a correção é uma noção não normativa aparentada à verdade: as emoções apresentam as coisas como sendo de uma certa maneira e elas são corretas quando as coisas são de fato dessa maneira (*vide* Seção 5 para mais debate). Tappolet contrasta este tipo de *neossentimentalismo representacional* com o neossentimentalismo normativo padrão. Enquanto tal, o primeiro tem ambições reducionistas – ele relaciona propriedades de valor às respostas corretas – mas não sustenta que Pierre é admirável *porque* é correto admirá-lo. Na verdade, o que ocorre é o contrário, sendo esta a razão pela qual Tappolet pensa que o neossentimentalismo representacional é naturalmente emparelhado com o realismo quanto ao valor (*vide* ODDIE, 2016). Enquanto os neossentimentalistas normativos

tomam as razões para valorizar como sendo fornecidas pelas propriedades não avaliativas do alvo (o que deixa em aberto a possibilidade de reduzir o valor às atitudes apropriadas), para Tappolet as razões para as atitudes avaliativas são fornecidas pelos próprios valores. A questão proeminente para a estratégia neossentimentalistas padrão é explicar a adequação ou as razões sem assumir valores independentes da resposta. O neossentimentalismo representacional, por outro lado, deixa para as emoções um papel primariamente epistêmico – como Tappolet encara isto, parte do que é possuir conceitos avaliativos é perceber que eles são canonicamente aplicados e justificados baseados em respostas emocionais (2016, p. 99). Confira o debate, novamente, na próxima seção.

É interessante que Adam Smith pode ser visto como oferecendo um tipo de explicação naturalista e sentimentalista da adequação ao criticar a análise intuicionista AA de seu tempo:

> [H]á alguns sistemas modernos de acordo com os quais a virtude consista na propriedade; ou na adequação da afeição [*affection*] a partir da qual agimos para com a causa ou objeto que a suscita [...] Nenhum desses sistemas fornecem, ou mesmo fingem fornecer, qualquer mensuração precisa ou distinta através da qual a adequação ou propriedade da afeição possa ser asserida ou julgada. Esta mensuração precisa e distinta só pode ser encontrada nos sentimentos simpáticos do espectador imparcial e bem informado (TSM, p. 346).

Pode-se naturalmente tomar isto, por exemplo, como a aceitação, por parte de Smith, da análise AA que a ação é errada se e somente se é apropriado ressentir-se com o agente por isso. Mas então ele fornece um molde distintamente sentimentalista para o que é para algo ser apropriado. Este fato normativo consiste na resposta sendo endossada por um espectador imparcial e bem informado (KAUPPINEN, 2014b).

5. Sentimentalismo Epistêmico

Sentimentalistas epistêmicos sustentam que (algumas) emoções são fontes do conhecimento moral ou ao menos da justificação moral. Muitos deles creem na

tese mais forte que experiências emocionais são necessárias para vir a conhecer verdades morais básicas, seja porque elas são fontes de justificação fundacional ou porque crenças morais justificadas devem ser coerentes com certas experiências emocionais. Logo, os sentimentalistas epistêmicos rejeitam tanto as visões intuicionistas, segundo as quais o entendimento adequado de proposições morais autoevidentes são suficientes para a crença justificada (AUDI, 2013), como as explicações racionalistas, segundo as quais a razão prática pura pode gerar conhecimento moral. Na base estão argumentos metafísicos e explicativos que mostram que verdades morais não são o tipo de relações ou fatos que o raciocínio pode descobrir (*vide* Suplemento **Antirracionalismo**). Note que isto não é negar que podemos raciocinar **a partir** de verdades morais básicas para verdades morais derivadas, ou que o raciocínio não moral não pode exercer um papel no conhecimento moral. Nem os sentimentalistas precisam sustentar que o conhecimento de verdades morais formais, como a transitividade do 'ser melhor que', é fundamentado em emoções.

Enquanto o sentimentalismo epistêmico tem, recentemente, ganhado popularidade, é bom ter em mente que sua principal tese é um tanto contraintuitiva. É um pensamento comum que as emoções podem nos desnortear em questões morais e prudenciais. Nós frequentemente passamos a perceber, retrospectivamente, que compreendemos mal uma situação como resultado de enxergá-la pelas lentes do ciúme, medo ou raiva. Não é apenas que emoções podem nos levar a agir contra os nossos melhores julgamentos, mas que elas podem levar crenças avaliativas falsas e injustificadas.

Os filósofos também frequentemente têm visto as emoções como epistemicamente problemáticas. Um pensamento muito antigo que aparece ao menos desde Platão é que as emoções são **disruptivas** ao prevenir e perverter a consideração racional. Para ter este tipo de efeito, as emoções não precisam ter qualquer conteúdo cognitivo – elas poderiam ser tão somente alterações mentais ou corporais.

Críticas mais interessantes da emoção garantem que elas têm conteúdo cognitivo, mas sustentam, não obstante, que elas são aptas a nos **enganar**. Na tradição ocidental, talvez sejam os estoicos os mais famosos por condenar emoções comuns por esta razão. Como base disto estava sua teoria do valor, segundo a qual apena a virtude é boa (a penas o vício é mau), enquanto coisas mundanas como a doença ou o sucesso social são indiferentes (LONG; SEDLEY, 1987). Uma vez que para eles nossas emoções ordinárias, da tristeza para a raiva e amor, são ou envolvem juízos quanto à maldade ou bondade das coisas mundanas que são realmente indiferentes, elas são **juízos avaliativos falsos**, ou impressões para

concordar com juízos falsos. Um quadro estruturalmente similar é oferecido pelos budistas que vêem especialmente emoções negativas como surgindo dos desejos e vínculos com coisas que não tem valor genuíno (FLANAGAN, 2016).

Dadas essas preocupações com disrupções e enganos, é justo dizer que os sentimentalistas epistêmicos têm uma presunção a superar. As próximas seções examinam algumas tentativas de fazer isto. A terceira subseção debate o papel potencial dos sentimentos nos contextos epistêmicos diferentes do da justificação básica, em particular nos casos de entendimento moral e a possibilidade de adquirir conhecimento moral por testemunho.

5.1 O Modelo Perceptual

Falar de um sentido moral sugere um quadro epistêmico segundo o qual o conhecimento moral é similar ao conhecimento adquirido pelos outros sentidos. Nós simplesmente prestamos atenção a algo real ou imaginado e, imediatamente, sem qualquer tipo de raciocínio, simplesmente sabemos se isto é certo ou errado, assim como passamos a saber da cor ou forma de um carro apenas olhando para ele. E sem dúvida é o caso que a crença na presença da propriedade natural fazedora do erro [*wrong-making*] (como causar intencionalmente uma dor intensa a um cachorro choramingando) desencadeia uma reação emocional em nós. É plausível, bem como apoiado nos dados empíricos (*vide* Seção 2.1), que a resposta emocional frequentemente precede e, talvez, cause o juízo moral. Mas seria isto uma percepção ou algo análogo?

No caso da experiência perceptual comum, um objeto instancia uma propriedade observável, tal como a de ser vermelho, e isto nos causa (da maneira correta) a correspondente representação fenomenal do objeto, tal como a experiência visual de um objeto vermelho na nossa frente (AUDI, 2013). Muitos creem que esta experiência não é ela mesma uma crença: eu posso crer que uma vara na água é reta e mesmo assim ela perceptualmente me aparece como não retilínea. Contudo, comumente ter uma experiência perceptual dá uma justificação para uma crença correspondente, seja porque ela apresenta ou representa as coisas como sendo de uma maneira particular (como os internistas epistêmicos pensam) ou porque ela confiavelmente nos informa que essas coisas são desta maneira (como os externistas epistêmicos pensam), ou uma combinação de ambos. Os epistemólogos debatem se a justificação requer apenas a ausência do derrotador [*defeater*] ou também de

algum tipo de garantia [*warrant*] para excluir cenários céticos, e o que precisar ser adicionado para uma crença verdadeira constituir **conhecimento**.

Assim, há algo como percepção moral emocional? A resposta depende, em parte, da natureza das emoções. Na visão de Hume, uma paixão é "uma existência original" que "não contém nenhuma qualidade representativa que a torne uma cópia de qualquer outra existência" (T, p. 415). Se este é o caso, então a experiência emocional não pode justificar a crença da mesma maneira que experiências perceptuais com conteúdo o podem. Contudo a maioria dos sentimentalistas epistêmicos contemporâneos rejeita a teoria das emoções de Hume. Em vez disto, eles subscrevem a alguma variedade de uma teoria representacional das emoções, motivados pelo pensamento que emoções podem ser avaliadas como adequadas ou apropriada ou mesmo racionais (DE SOUSA, 1987). Isto sugere que emoções são **sobre** algo – que elas têm conteúdo intencional.

Uma teoria representacional sustenta que emoções envolvem ou consistem em um **julgamento** que o alvo da emoção, tal como voar em um avião, tem a propriedade que é o **objeto formal** da emoção, tal como ser teimoso ou perigoso no caso do medo (KENNY, 1963; LYONS, 1980). Sabine Roeser (2011) defende este tipo de concepção no caso da moral. Segundo ela, emoções morais "são ao mesmo tempo tanto juízos de valor como estados afetivos" (ROESER, 2011, p. 149). Na sua visão, elas constituem intuições no mesmo sentido que intuicionistas tradicionais sustentam: elas são apreensões diretas de estados de coisas não naturais (este é um caso evidente no qual as variantes epistêmicas e metafísicas do sentimentalismo se separam). Qualquer visão deste tipo encara o desafio de explicar por que emoções iriam confiavelmente rastrear tais fatos não naturais. A outra preocupação é que juízos de valor não são, em geral, autojustificáveis, de modo que não é claro porque juízos embutidos na emoção iriam justificar crenças ulteriores.

De todo modo, trata-se de um problema bem conhecido com teorias de julgamento da emoção que emoções podem ser **recalcitrantes**, isto é, se separar dos nossos julgamentos. Nós podemos estar preocupados apesar de julgar que tudo está bem. Ademais, muitos pensam que as crianças ou animais podem ter emoções sem possuir conceitos avaliativos e, assim, sem serem capazes de fazer julgamentos avaliativos. Estas considerações levaram muitos a considerarem as emoções como sendo, ou contendo, ou sendo análogas às *experiências perceptuais* (DE SOUSA, 1987; TAPPOLET, 2000, 2011; HELM, 2001; PRINZ, 2004, 2007; ZAGZEBSKI, 2003; GOLDIE, 2007; DÖRING, 2007; WISNEWSKI, 2015). Neste tipo de visão, as emoções e as experiências perceptuais têm muitas características

em comum: elas são espontâneas (não estando sob controle direto da volição), informativamente encapsuladas (isto é, relativamente independentes das crenças e desejos do sujeito), têm uma característica fenomenológica e têm conteúdo conceitual (e/ou não conceitual), assim representando ou construindo o mundo como sendo de uma certa maneira sem envolver julgamento ou crença.

Nas visões perceptualistas, as emoções são as coisas mais próximas que temos do contato [*acquaintance*] com o valor – por exemplo, como Adam Pelser coloca, "experenciar algo como divertido é simplesmente se divertir com isto" (2014, p. 112). Assim como pessoas daltônicas podem conhecer na base do testemunho que um objeto é vermelho sem estar em contato com o vermelho, uma pessoa que é incapaz de se espantar [*awe*] pode aprender com os outros que algo é sublime, enquanto é desprovida do contato com ou acesso direto a sua sublimidade (*ibid*.). Alegações análogas têm também sido defendidas quanto à natureza do *desejo*, que é sustentado por muitos como sendo uma aparência não doxástica do bem (STAMPE, 1987; ODDIE, 2005; TENENBAUM, 2007; SINHABABU, 2017). Perceptualistas tendem a sustentar que a característica distintiva da emoção perceptual é que o conteúdo é avaliativo, o que por sua vez é frequentemente pensado como explicando a fenomenologia especial. Contudo, Julien Deonna e Fabrice Teroni (2015) têm recentemente argumentado que o caráter avaliativo das emoções deve ser compreendido em termos de *atitudes* corporais, não conteúdo avaliativo (*vide* MITCHELL, no prelo). Para mais debates acerca da natureza da emoção, confira o verbete sobre **emoção**.

Dada a tese metafísica perceptualista quanto às emoções, muitos têm argumentado para o seguinte tipo de tese epistêmica:

> **Perceptualismo Emocional Epistêmico (PEE):** Emoções são as fontes de justificação imediata (não inferencial) prima facie (ou derrotável) para proposições avaliativas (*vide* CARTER, 2019).

Se o Perceptualismo Emocional Epistêmico é verdadeiro, as emoções podem desempenhar um papel fundacional na epistemologia moral – elas (ou seus conteúdos) podem potencialmente justificar a crença em proposições morais independentemente do conteúdo das outras crenças da pessoa, assim como minha experiência perceptual de um cachorro preto na minha frente pode potencialmente justificar a crença na proposição correspondente, independentemente do que eu

acreditava antes (*vide* DÖRING, 2007; PELSER, 2014). Certamente é possível assumir uma posição de compromisso, de acordo com a qual justificação fornecida pela emoção é suficiente para justificar a crença apenas se ela também é coerente com as outras crenças da pessoa, como nota van Roojen (2014).

Contudo, como Robert Cowan (2016) enfatiza, o PEE não é implicado pelo perceptualismo quanto à emoção, dado que as diferenças metafísicas entre as emoções e as experiências perceptuais paradigmáticas podem contribuir para uma diferença epistêmica, como de fato muitos têm argumentado.

Assim, vale a pena perguntar por que as emoções seriam fontes de justificação, mesmo se elas forem análogas à, ou uma espécie distinta de, experiência perceptual. Christine Toppolet argumenta que já que a justificação prima facie é um status epistêmico mínimo, é fácil para obter: "Se você está com medo de um cachorro, é certamente algo que torna sua crença que o cachorro é temível prima facie justificada." (TOPPOLET, 2016, p. 40). Contudo, se formos estar justificados em acreditar na base da experiência emocional, qualquer que seja a justificação que ela ofereça deve ser forte o suficiente para a crença. Alguma explicação para isto, baseada em princípios, é necessária. A maneira mais simples seria sustentar que qualquer aparência das coisas sendo de uma certa maneira pode gerar uma justificação suficiente, ainda que derrotável, tal como o *liberalismo epistêmico*, de Michael Huemer (2001) sustenta. Tanto quanto o medo, por exemplo, apresenta o seu alvo como sendo temível (ou perigoso), ele justificaria a crença avaliativa correspondente na ausência de derrotadores (*vide* TUCKER, 2013).

Assim, enquanto há um esquema epistêmico plausível (mesmo que controverso) que explicaria o PEE, há muitos desafios sérios contra ele. Afinal, mesmo proponentes da abordagem perceptualista da emoção concedem que haja uma série de diferenças notáveis entre emoções e experiências sensoriais perceptuais (TAPPOLET, 2016). Algumas dessas diferenças podem ser epistemicamente relevantes. Para começar, parece haver uma diferença óbvia na nossa prática epistêmica cotidiana: tipicamente nós não pedimos razões para experiências perceptuais, mas pedimos razões para emoções (BRADY, 2013). Por exemplo, se você diz que vê um cachorro lá fora, é improvável que eu pergunte "Por quê?", enquanto se você diz que está com medo do cachorro, faz perfeitamente sentido fazer esta questão. Inversamente, se eu te pergunto por que a ponte é perigosa, provavelmente você não dirá "Porque eu estou com medo dela", enquanto você pode apelar para ter visto o cachorro lá fora para justificar sua crença que ele está

na ponte. Como Michael Brady diz, diferentemente de percepções sensoriais, as emoções suscitam demandas justificativas (2013, p. 110), em vez de silenciá-las. Ademais, se uma emoção como o medo justifica crenças quanto à periculosidade, é bem questionável que isto seria autojustificativo, dado que o perigo merece o medo e "O próprio fato que estou com medo do cachorro não pode, por si só, ser uma evidência que é apropriado ou adequado ter medo do cachorro" (BRADY, 2010, p. 123).

Como os perceptualistas epistêmicos poderiam responder? Uma possibilidade é enfatizar uma diferença saliente entre as emoções e as percepções sensoriais: emoções têm **bases cognitivas** nas percepções dos sujeitos ou nas crenças (DEONNA; TERONI, 2012, p. 24-25). Por exemplo, eu não tenho medo do cachorro à minha frente a menos que haja um cachorro à minha frente e eu creia que ele possa me machucar. Meu medo é fundamentado em parte nessas crenças e percepções. Uma consequência é que a própria emoção é justificada apenas se as bases cognitivas são justificadas (ainda que não necessariamente verdadeiras) (GREENSPAN, 1988). Se assim o for, a justificação emocional é **dependente** (PELSER, 2014; COWAN, 2018). Inicialmente, isto pode parecer uma má notícia para o PEE, já que ela parece conflitar com a tese que emoções fornecem justificação imediata para proposições avaliativas. Mas note que enquanto a justificação emocional pode pressupor a justificação para a crença em **proposições não avaliativas** (tais como aquelas que dizem respeito à propensão de um cachorro para atacar), há uma lacuna entre estes ***inputs*** e a proposição **avaliativa** que é o conteúdo da própria experiência emocional, de acordo com o perceptualismo (*vide* ETCHEVERRI, 2019). Eu posso estar com raiva de você porque eu acredito que você contou a alguém o meu hobby peculiar, mas a raiva em si mesma não te apresenta como tendo contado a respeito do meu hobby, mas, em vez disso, como tendo agido mal comigo. Mesmo que seja uma questão de fato, o caso de terem agido mal comigo nesta situação consiste em você ter revelado um fato embaraçoso a meu respeito, mas disto não se segue que aquilo que justifica a crença na proposição factual é suficiente para justificar a crença na proposição moral. Por tudo que foi dito quanto à dependência epistêmica com base cognitiva, minha raiva pode fornecer justificação imediata e **adicional** par a crença quanto ao erro da sua ação. Como Robert Cowan coloca isto, "quando a base cognitiva de uma emoção é justificada, então as emoções podem justificar crenças com conteúdo avaliativo, independentemente de terem justificação ulterior para proposições avaliativas apoiadoras de crenças" (2018, p. 229).

O lado positivo da dependência epistêmica é, então, que ela permite aos perceptualistas epistêmicos darem uma resposta natural aos desafios das questões

por quê. Se você me pergunta por que eu penso que o lobo é perigoso ou por que eu penso que a ação de meu amigo é errada, eu irei citar a característica não avaliativa saliente que torna a situação perigosa ou o ato errado e, assim, justificar minha emoções (como eu vejo as coisas), em vez do fato que eu respondo a isto com uma emoção. Pelas lentes da minha emoção, por assim dizer, eu tomo esta característica não avaliativa como fundamentando uma verdade avaliativa a respeito do alvo – irá parecer óbvio para mim que não vale a pena mencionar que, digamos, um cachorro estando pronto para atacar é digno de medo.

Mas o que dizer do desafio da autojustificação? Jonathan Mitchell (2017) assinala que em casos não patológicos, o medo, por exemplo, não apresenta o seu alvo como chamando o medo – de outra maneira, a emoção não é inteligível para o próprio sujeito. Plausivelmente Brady está correto que a evidência para o perigo (ser digno de medo) não é o fato que estou com medo, mas o cachorro constituir uma ameaça para os meus interesses. Não obstante, o meu medo pode fazer o status de fornecedor de razão [*reason-giving*] deste fato **avaliável** para mim no caso bom. O medo ainda não seria autojustificativo, visto que o cacho pode, de fato, não constituir uma ameaça.

Entretanto, os problemas para o perceptualismo não terminam com a dependência a partir de bases cognitivas. Há uma razão para acreditar que as emoções também dependem dos nossos **desejos** ou preocupações (ROBERTS, 2003). Eu só fico triste quando você quebra um vaso antigo quando eu gostaria de tê-lo ou só tenho medo do lobo rosnando se eu não quero ser comido. Quando se trata de crença ou mesmo experiência perceptual, é um golpe contra a posição da justificação se o estado é influenciado por um desejo – como Markie (2005) observa ao criticar o liberalismo epistêmico, se uma pedra parece uma pepita de ouro apenas porque você realmente quer que ela seja uma pepita de ouro, este aparecer não justifica a crença que ela é uma pepita de ouro. Assim, por que este tipo de penetrabilidade cognitiva não é um problema para o PEE? Uma maneira natural de contornar o problema é dizer que os desejos ou preocupações subjacentes à aparência emocional devem eles mesmos serem corretos ou justificados. O problema com essa sugestão é que então se necessita de uma história para como os desejos são corretos ou justificados que não apele, ela própria, para a experiência emocional, o que não parece ser compatível com as emoções desempenhando um papel epistêmico fundacional (*vide* HELM, 2001).

Por fim, enquanto o PEE tende a ter um caráter epistêmico internista, a experiência perceptual em geral é evidentemente ligada ao mundo externo. No caso

da percepção sensorial, os elos causais com os objetos da percepção desempenham um papel em explicar a confiabilidade e, pode-se argumentar, o status justificativo, das experiências perceptuais. Contudo, as emoções não parecem ser causadas pelas propriedades avaliativas que elas representam – embora se nós considerarmos como vamos dos estímulos próximos sensoriais para os objetos da percepção, a diferença entre experiências perceptuais comuns e as emoções pode não ser tão grande (MILONA; NAAR, 2020). As relações causais podem também explicar por que as experiências perceptuais possuem o **conteúdo** que elas possuem.

Uma linha alternativa é dizer que o que causa a emoção não é uma característica avaliativa do objeto – em vez disso, o que nós literalmente percebemos é a propriedade que fundamenta o erro da ação, tal como o policial atirar em um manifestante pacífico, e então responder a isto com uma reação emocional, o que constrói a ação como errada. A imaginação também funcionaria igualmente bem. Sendo assim, talvez seja melhor pensar emoções como aparências quase-perceptuais que constituem intuições morais, não percepções (KAUPPINEN, 2013). Isso se relaciona naturalmente com uma visão na qual emoções (re)presentam seus alvos como tendo certas propriedades avaliativas em virtude de seu caráter fenomenal, não de sua etiologia. A maneira que se sente uma emoção não é irrelevante para como ela constrói o seu objeto. Confira Goldie (2000) sobre "sentimentos direcionados" a um objeto. Os sentimentos envolvidos na raiva, por exemplo, não são sensações sem direção, mas tem um caráter qualitativo em virtude do qual a raiva apresenta seu objeto como ofensivo (KRIEGEL, 2014). E o que dizer da confiabilidade? Novamente, muitos acreditam que a experiência perceptual comum justifica a crença, em parte porque ela é em geral confiável, ainda que suas características fenomenais sejam também relevantes para a justificação. Em todo caso, o **conhecimento** moral requer mais do que justificação e verdade – como é comum dizer atualmente, uma crença deve ser **segura** (tal que não poderia facilmente ter sido o caso que a pessoa teria formado uma crença falsa usando o mesmo método) para constituir conhecimento (SOSA, 1999; WILLIAMSON, 2000). Como observado anteriormente, alguns sentimentalistas epistêmicos, como Tappolet (2011), não são sentimentalistas metafísicos, mas realistas quanto ao valor. Aqueles que sustentam uma visão deste tipo podem argumentar, como o faz Adam Carter (2019), que para as experiências emocionais justificarem uma crença, elas devem surgir de uma **competência** ou habilidade, grosso modo uma disposição para acertar [*get it right*] quando apropriadamente situadas. Outros perceptualistas epistêmicos, como Prinz (2007), subscrevem a uma metafísica sentimentalista. Há espaço para eles oferecerem

uma **explicação constitutiva** da confiabilidade das emoções: posto que fatos avaliativos são constitutivamente ligados às respostas emocionais, há uma boa razão para pensar que as respostas emocionais os rastreiam, ao menos sob condições apropriadas.

Certamente, mesmo se a percepção moral (ou intuição sentimental) for epistemicamente tão boa como a percepção sensorial, nós sabemos que em algumas circunstâncias nossos sentidos não são confiáveis. Mas como sabemos que circunstâncias são essas? Este **desafio da calibragem** já foi notado por Hutcheson (1728, p. 177):

> Mas não poderia haver um estado certo ou errado do nosso senso moral, tal como há nos nossos outros sentidos, conforme eles representam seus objetos para serem como eles realmente são, ou representá-los de outra maneira? Então não poderia o nosso senso moral aprovar aquilo que é vicioso e desaprovar a virtude, tal como um paladar doentio poderia não gostar de comida agradável, ou uma visão viciada deturpar as cores e as dimensões? Portanto, não deveríamos saber o que é moralmente bom ou mau através de nossa razão, antes de podermos saber que o nosso senso moral está certo?

A resposta de Hutcheson era para assinalar que ainda que o senso moral possa enganar, disto não se segue que a **razão** é necessária para fornecer um padrão para isto, ou que é capaz de fazer isto. Afinal, não é através do raciocínio que nós fundamentalmente corrigimos os equívocos da percepção de cor. Em vez disto, tentamos determinar quais circunstâncias conduzem à percepção que funciona apropriadamente (talvez através de uma investigação *a priori*) e então tentar nos colocar nessas circunstâncias e olhar novamente. Como são as circunstâncias apropriadas nos casos morais é algo que depende da natureza dos fatos morais. Para relativistas como Prinz (2007), a única coisa que importa é que nossas reações genuinamente reflitam nossos sentimentos quanto à ação. Para disposicionalistas ideais, David Lewis (1989, p. 117) oferece uma receita básica:

> Para descobrir se estaríamos dispostos, sob condições ideias, a valorizar isto, se coloque em condições

> ideias, se você puder, certificando-se que você pode determinar quando você foi bem sucedido em fazê-lo. Então descubra se você valoriza a coisa em questão, i.e., se você tem o desejo de desejá-la. Se você tem, isto confirma que isto é um valor.

Para o próprio Lewis, as condições ideias para valorizar são aquelas de contato [*acquaintance*] imaginativo completo com o objeto. Para Hume e Smith, as condições ideias para o juízo moral envolvem ocupar o "ponto de vista comum", ou a perspectiva de um espectador imparcial e bem informado. Como debatido na seção 2.2, ambos creem que atingir isto requer simpatizar (ou ter empatia) com os sentimentos ou atitudes reativas daqueles imediatamente afetados, enquanto se controla as distorções previsíveis. A menos que elas sejam sentidas a partir do ponto de vista comum, nossas repostas emocionais não rastrearão confiavelmente os fatos morais dependentes de resposta.

5.2. Teorias do Rastreamento

Nem todos os sentimentalistas epistêmicos subscrevem ao modelo perceptualista. Uma alternativa é fornecida por John Allman e Jim Woodward (2008), que argumentam que intuições morais (no sentido de juízos imediatos resultantes do que os psicólogos chamam de Sistema 1) envolvendo processamentos emocionais podem confiavelmente rastrear fatos morais depois de um tipo apropriado de aprendizado implícito. Aqui eles recorrem ao trabalho empírico quanto ao desenvolvimento de *expertise*. No aprendizado implícito, o aprendiz, tal como uma enfermeira novata, é repetidamente exposto a certas pistas (tal como bebês manifestando diferentes sintomas), tomam certas decisões para agir (dando certo tipo de tratamento para o bebê) e então recebem um retorno claro e independente quanto a se as decisões foram corretas (os sintomas ou vão embora, ou pioram) (KAHNEMAN; KLEIN, 2009). Como resultado, o aprendiz pode se tornar um expert, sendo capaz de responder imediatamente e apropriadamente às situações sem necessariamente ser capaz de articular o porquê. A sugestão de Almann e Woodward é que um tipo similar de treinamento de nossas respostas morais é possível no caso da moral.

Mais recentemente, Peter Railton também sublinhou concepções do sistema

afetivo como "um sistema processador de informação, baseado na experiência e flexível, muito capaz de rastrear dependências estatísticas e de guiar a seleção de comportamento via o balanceamento de custos, benefícios e riscos (2014, p. 833). Assim como tem sido descoberto que ratos de laboratórios desenvolvem modelos avançados de direcionamento de ação os quais implicitamente representam os valores esperados de suas opções, os nossos sentimentos podem ser sintonizados às possíveis recompensas e punições, podendo-se esperar que eles deem o tom [*to color*] da nossa percepção avaliativa.

O principal desafio para este tipo de visão é que há uma diferença entre, por um lado, treinar as emoções para responderem às informações potencialmente **moralmente relevantes**, tais como

> os estados mentais dos outros e seus prováveis comportamentos em situações interativas, assim como as consequências prováveis deste comportamento (ALLMANN; WOODWARD, 2008, p. 173).

e, por outro lado, treinar as emoções para responderem aos próprios **fatos morais**. Não é claro se há qualquer analogia entre o retorno independente e não ambíguo que uma enfermeira recebe depois de dar o tratamento errado e o caso de fazer um juízo moral equivocado. Railton sugere que a empatia pode desempenhar este papel, observado que ter uma capacidade de ter empatia e ativar *in situ* "(parecem [...] ser importantes preditores do comportamento e da sensibilidade moral" (2014, p. 844), por exemplo, ao inibir a violência. Mas enquanto a empatia pode certamente dar um retorno quanto às respostas (antecipadas) dos outros, uma coisa é aprender que os outros, como uma questão de fato, digamos, sofrem por ou desaprovam o que eu fiz, outra coisa bem diferente é aprender que esta desaprovação é **merecida**. Parece que é este último tipo de retorno que nós precisaríamos para sintonizar respostas afetivas aos fatos morais.

5.3. Entendimento Moral e Testemunho

Se ou não as emoções são a fonte da justificação avaliativa ou do conhecimento, elas podem desempenhar um papel quanto a um bem epistêmico diferente, a saber, o **entendimento moral**, e, talvez de maneira relacionada, ajudar

a explicar por que nossas práticas morais epistêmicas são diferentes de muitas outras práticas morais. Começando pela última, considere os seguintes casos de testemunho (*vide* CALLAHAN, 2018):

> Nina quer saber se a lei permite deduzir despesas decorrentes de trabalhos de caridade ao declarar seus impostos. Ela pergunta a um especialista, que lhe diz que sim. Assim, Nina passa a acreditar que a lei permite deduzir as despesas.
>
> Aleksi quer saber se é, enquanto uma regra, moralmente errado enganar os outros ao omitir detalhes desfavoráveis. Ele pergunta a um amigo sábio, que diz que sim. Aleksi passa, assim, a acreditar que é, enquanto uma regra moralmente errada, mentir por omissão.

Vamos estipular que ambos, Nina e Aleksi, passam a saber aquilo que eles gostariam de saber pelo testemunho (deixe de lado qualquer preocupação geral quanto ao conhecimento testemunhal e assume que em ambos os casos eles tem razão suficiente para confiar na palavra da pessoa que eles estão perguntando). Para muitos parece que o que Nina faz é inteiramente não problemático, mas há algo de errado, se não completamente errado, com a **deferência moral** de Aleksi. De maneira geral, uma qualificação comumente aceita é que a deferência moral é problemática quando se forma uma crença baseada tão somente no testemunho, em vez de, por exemplo, refletir sobre a questão, como resultado do outro expressar uma opinião. Outra qualificação comum é que o testemunho problemático é aquele acerca de um conteúdo moral puro cuja verdade não depende de suposições empíricas. Não é problemático mudar sua opinião quanto a uma questão moral impura tomando como base um testemunho que diz respeito a um fato empírico moralmente relevante, como quando alguém passa a crer que é errado comer peixe com base em ser informado que peixes sentem dor (FLETCHER, 2016). Em suma, parece que você não deve mudar sua mente, quanto a uma questão puramente moral, apenas porque uma pessoa te diz para fazê-lo (mesmo que você não tenha nenhuma razão particular para duvidar dela).

Mas, se é que há algo, o que há de errado com a deferência moral? Guy Fletcher argumenta que a ligação entre sentimentos e moralidade explica este desconforto. Primeiro, ele sustenta que "É difícil formar sentimentos morais tomando

como base testemunho puro e direto (FLETCHER, 2016, p. 60). Isto encontra paralelo em outros casos: por exemplo, eu não posso sentir admiração por uma pintura só porque você me diz que vale a pena admirá-la. Segundo, como debatido na seção 2, há várias visões metaéticas nas quais juízos morais consistem (parcialmente ou completamente) em sentimentos ou são ligados aos sentimentos. Se for assim, juízos baseados no testemunho são impossíveis ou deficientes. Contudo, Laura Frances Callahan (2018) assinala que se alguém **formou** sentimentos morais puramente na base do testemunho, nossa sensação de que algo suspeito está acontecendo não iria desaparecer. De fato, seria ainda mais alarmante se alguém se tornasse fervoroso quanto a alguma questão moral apenas porque outra pessoa o disse que isto era errado!

Uma explicação popular muito diferente quanto à natureza problemática da deferência moral apela para a noção de entendimento moral: enquanto podemos adquirir conhecimento **moral** pelo testemunho, nós não iremos (facilmente ou de todo) adquirir **compreensão** moral desta maneira, mas a compreensão moral é importante para o valor moral ou para merecer crédito quanto a sua ação (HILLS, 2009). Para Alison Hills (2016), a compreensão moral é uma forma de "controle cognitivo", grosso modo a habilidade de captar e articular porque certas características fazem de uma ação (digamos) incorreta e, consequentemente, o que teria sido necessário ser diferente a fim de que ela fosse correta. Callahan (2018) argumenta que ter o tipo de habilidade cognitiva que Hills destaca não é suficiente para o valor moral, já que as possuir é consistente com estar motivado por características que não são razões para ação. Na sua abordagem sentimentalista alternativa, a própria compreensão moral requer também ter respostas motivacionais, e afetivamente apropriadas, para as proposições compreendidas e as relações explicativas relevantes (CALLAHAN, 2018, p. 451-452). Por exemplo, compreender que uma prática é errada porque ela é racista envolve uma disposição de ter raiva ou ficar triste com ela porque ela é racista (ou quanto às características que a torna racista) e uma vontade de se abster e de intervir nesta prática. Assim, o problema com a deferência é que ela desincentiva o tipo de reflexão e investigação que suscitariam estas respostas, além das habilidades cognitivas relevantes (CALLAHAN, 2018, p. 454-456). Para uma abordagem sentimentalista muito diferente, confira Howard (2018). Por fim, mesmo aqueles que não acreditam que sentimentos são constitutivos da compreensão moral podem acreditar que eles ajudam a suscitá-lo. Assim, Brady (2013) argumenta que o papel das emoções em nos **motivar** a descobrir o porquê de algumas coisas serem certas ou erradas constitui em grande parte o papel epistêmico positivo das emoções na moralidade.

6. Uma Variedade de Sentimentalismos Morais (Reprise)

No início notamos que há muitas coisas sobre as quais se pode ser um sentimentalista. Agora estamos em posição de fazer algumas observações quanto às relações entre elas. Primeiro, o *sentimentalismo explicativo* é claramente independente dos outros tipos. É perfeitamente possível que emoções desempenhem algum papel significativo na explicação do juízo moral, independente do que consistam este juízos ou fatos morais e se eles são ou não fontes do conhecimento moral. Se alguma forma simples do sentimentalismo de juízo for verdadeira, certamente as emoções oferecem um tipo de explicação constitutiva dos juízos morais – se eu estou com raiva de você, isto explica por que eu moralmente julgo que você errou comigo, porque isto constitui o meu juízo. Ademais, o sentimentalismo explicativo claramente tem implicações epistêmicas. Se as emoções tanto explicam como justificam as nossas crenças, nossos vereditos morais são vindicados. Se, em contraste, as emoções explicam nossas crenças, mas **não** as justificam – se o sentimentalismo explicativo é verdadeiro, mas o sentimentalismo epistêmico é falso – nós caímos em uma visão cética.

Segundo, o *sentimentalismo de juízo* tem um parentesco evidente com o sentimentalismo **metafísico**. É assim especialmente com aquelas versões cognitivistas segundo as quais nossos juízos atribuem propriedades dependentes de reposta, as quais são metafisicamente ligadas as nossas respostas sentimentais. Mas há também contextualistas para os quais os juízos morais descrevem propriedades naturais comuns, e o os teóricos das atitudes apropriadas, para os quais os juízos morais são sobre a adequação das atitudes, onde fatos sobre a adequação podem eles mesmos ser verdades não naturais.

Por fim, o *sentimentalismo epistêmico* é novamente, em princípio, independente das outras variedades: nós vimos que alguém pode ser um realista independente da mente quanto às propriedades morais, enquanto pensa que algumas emoções apropriadas fornecem acesso às verdades morais. Mas se fatos morais são constitutivamente relacionados às nossas reações e os nossos juízos são sobre estes fatos, teremos uma boa razão *a priori* para pensar que crenças baseadas em emoções apropriadas não iriam facilmente nos enganar e podem, assim, equivaler ao conhecimento. Assim, enquanto formas diferentes de sentimentalismo são separáveis umas das outras, elas se apoiam mutuamente, se mantidas em conjunto.

Referência bibliográfica

ALTOLA, E., 2014, "Affective Empathy as Core Moral Agency: Psychopathy, Autism and Reason Revisited," *Philosophical Explorations*, 17 (1): 76-92.

ALLMAN, J. and J. Woodward, 2008, "What Are Moral Intuitions and Why Should We Care About Them? A Neurobiological Perspective," *Philosophical Issues*, 18: 164-185.

AUDI, R., 2013, *Moral Perception*, Princeton, NJ: Princeton University Press.

AYER, A.J., 1936, *Language, Truth and Logic*, London: Gollancz.

BAGNOLI, C. (ed.), 2011, *Morality and the Emotions*, Oxford: Oxford University Press.

BARKER, S., 2000, "Is Value Content a Component of Conventional Implicature?" *Analysis*, 60: 268-279.

BISHOP, J. D., 1996, "Moral Motivation and the Development of Francis Hutcheson's Philosophy," *Journal of the History of Ideas*, 57 (2): 277-295.

BJÖRNSSON, G. and Finlay, S., 2010, "MetaethicalContextualism Defended," *Ethics*, 121(1): 7-36.

BLACKBURN, S., 1971, "Moral Realism", in J. Casey (ed.), *Morality and Moral Reasoning*, London: Methuen, pp. 101-124.

BLACKBURN, S., 1985, "Supervenience Revisited", in Ian Hacking (ed.), *Exercises in Analysis*, Cambridge: Cambridge University Press, pp. 47-68.

BLACKBURN, S., 1988, "How To Be an Ethical Anti-Realist," *Midwest Studies in Philosophy*, 12(1): 361-375.

BLACKBURN, S., 1993, "Circles, Finks, Smells, and Biconditionals," *Philosophical Perspectives*, 7: 259-279.

BLACKBURN, S., 1998, *Ruling Passions*, Oxford: Clarendon Press.

BLACKBURN, S., 2002, "Précis of Ruling Passions," *Philosophy and Phenomenological Research*, 65(1): 122-135.

BLOOM, P., 2017, *Against Empathy: The Case for Rational Compassion*, New York: Random House.

BOISVERT, D., 2008, "Expressive-Assertivism," *Pacific Philosophical Quarterly*, 89(2): 169-203.

BRADY, M. S., 2010, "Virtue, Emotion and Attention," *Metaphilosophy*, 41: 115-131.

BRADY, M. S., 2013, *Emotional Insight: The Epistemic Role of Emotional Experience*, Oxford: Oxford University Press.

BRANDT, R.B., 1955, "The Definition of an 'Ideal Observer' Theory in Ethics," *Philosophy and Phenomenological Research*, 15(3): 407-413.

BRINK, D., 1986, “Externalist Moral Realism,” *Southern Journal of Philosophy*, 24(Supplement): 23-42.
BRINK, D., 1989, *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
BROAD, C. D., 1944-45, “Some Reflections on Moral-Sense Theories in Ethics,” *Proceedings of the Aristotelian Society*, 45: 131-166.
CALLAHAN, L. F., 2018, “Moral Testimony: A Re-Conceived Understanding Explanation,” *Philosophical Quarterly*, 68 (272): 437-459.
CARTER, J. A., 2019, “Epistemic Perceptualism, Skill, and the Regress Problem,” *Philosophical Studies*, 1-26.
COHON, R., 2008, *Hume’s Morality: Feeling and Fabrication*, Oxford: Oxford University Press.
COWAN, R., 2018, “Epistemic Sentimentalism and Epistemic Reason-Responsiveness,” *Evaluative Perception*, 219.
CUNEO, T., 2001, “Are Moral Qualities Response-Dependent?” *Noûs*, 35(4): 569-591.
D’ARMS, J., 2011, “Empathy, Approval, and Disapproval in *Moral Sentimentalism*,” *Southern Journal of Philosophy*, 49(Supplement): 134-141.
D’ARMS, J. and D. JACOBSON, 2000a, “Sentiment and Value,” *Ethics*, 110(4): 722-48.
D’ARMS, J. and D. JACOBSON, 2000b, “The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions,” *Philosophy and Phenomenological Research*, 61: 65-90.
D’ARMS, J. and D. JACOBSON, 2006, “Sensibility Theory and Projectivism,” in *The Oxford Handbook of Ethical Theory*, D. Copp (ed.), Oxford: Oxford University Press, pp. 186-218.
DARWALL, S., 1998, “Empathy, Sympathy, and Care,” *Philosophical Studies*, 89: 261-282.
DEONNA, J. and TERONI, F., 2012, *The Emotions: A Philosophical Introduction*, London: Routledge.
DEONNA, J. and TERONI, F., 2015, “Emotions as Attitudes,” *Dialectica*69 (3):293-311.
DE SOUSA, R.B., 1987, *The Rationality of Emotion*, Cambridge, MA: MIT Press.
DE WAAL, F., 2009, *The Age of Empathy*, New York: Harmony Books.
DÖRING, S., 2007, “Seeing What to Do: Affective Perception and Rational Motivation,” *dialectica*, 61, 363-393.
DREIER, J., 1990, “Internalism and Speaker Relativism,” *Ethics*, 101: 6-26.
DREIER, J., 2009, “Relativism (and Expressivism) and the Problem of Disagreement,” *Philosophical Perspectives*, 23(1): 79-110.
DRIVER, J., 2011, “The Secret Chain: A Limited Defense of Sympathy,” *Southern Journal of Philosophy*, 49 (Supplement): 234-238.

EGAN, A., 2012, "Relativist Dispositional Theories of Value," *Southern Journal of Philosophy*, 50(4): 557-582.
ENOCH, D., 2005, "Why Idealize?" *Ethics*, 115(4): 759-787.
ECHEVERRI, S., 2019, "Emotional Justification," *Philosophy and Phenomenological Research*, 98 (3): 541-566.
EWING, A. C., 1948, *The Definition of Good*, London: Routledge&Kegan Paul.
FELDMAN, F., 2004, *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*, New York: Oxford University Press.
FINLAY, S., 2017, "Disagreement Lost and Found," *Oxford Studies in Metaethics,* 12: 187-205.
FIRTH, R., 1952, "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research*, 12: 317-345.
FLANAGAN, O., 2016, *The Geography of Morals: Varieties of Moral Possibility*, New York: Oxford University Press.
FLEISCHACKER, S., 2019, *Being Me Being You: Adam Smith and Empathy*, Chicago: University of Chicago Press.
FLETCHER, G., 2016, "Moral Testimony: Once More with Feeling," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 11*, Oxford University Press, pp. 45-73.
FLETCHER, G. and M. RIDGE, 2014, *Having It Both Ways: Hybrid Theories and Modern Metaethics*, Oxford University Press.
FROST, K., 2014, "On the Very Idea of Direction of Fit," *Philosophical Review*, 123 (4): 429-484.
GENDLER, T. S., 2008, "Alief in Action (and Reaction)," *Mind & Language*, 23 (5): 552-585.
GIBBARD, A., 1990, *Wise Choices, Apt Feelings*, Cambridge: Harvard University Press.
GIBBARD, A., 2003.*Thinking How to Live*, Cambridge: Harvard University Press.
GREENE, J.D., 2008, "The Secret Joke of Kant's Soul," in Sinnott-Armstrong (ed.) 2008, Vol. 3, pp. 35-80.
GREENE, J.D., R.B. SOMMERVILLE, L.E. NYSTROM, J.M. DARLEY, and J.D. COHEN, 2001, "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science*, 293: 2105-2108.
GREENE, J.D. AND J. HAIDT, 2002, "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences*, 6(12): 517-523.
GREENE, J.D., F.A. CUSHMAN, L.E. STEWART, K. LOWENBERG, L.E. NYSTROM, and J.D. COHEN, 2009, "Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment," *Cognition*, 111(3): 364-371.

GOLDIE, P., 2000, *The Emotions: A Philosophical Exploration*, Oxford: Oxford University Press.
GOLDIE, P., 2007, "Seeing What Is the Kind Thing to Do: Perception and Emotion in Morality," *dialectica*, 61(3): 347-361.
GORDON, R.M., 1995, "Sympathy, Simulation, and the Impartial Spectator," *Ethics*, 105(4): 727-742.
GREENSPAN, P., 1988, *Emotions and Reasons: An Enquiry into Emotional Justification*, Routledge.
GROUCHY, S. de, 1798, *Sophie de Grouchy's Letters on Sympathy*, translated by S. Bergès, with an introduction by S. Bergès and E. Schliesser, Oxford: Oxford University Press, 2019.
HAIDT, J., 2001, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review*, 108: 814-34.
HAIDT, J., 2012, *The Righteous Mind*, New York: Pantheon.
HAIDT, J. and F. BJÖRKLUND, 2008, "Social Intuitionists Answer Six Questions About Moral Psychology," in Sinnott-Armstrong (ed.) 2008, Vol. 2, pp. 181-217.
HARE, R.M., 1952, *The Language of Morals*, Oxford: Clarendon.
HARMAN, G., 1975, "Moral Relativism Defended," *Philosophical Review*, 84: 3-22.
HARMAN, G., 2000, "What Is Moral Relativism?" reprinted in *Explaining Value*, New York: Oxford University Press, 20-38.
HAYWARD, M. K., 2019, "Practical Reason, Sympathy and Reactive Attitudes," *Noûs*, 53 (1): 51-75.
HELM, B., 2001, *Emotional Reason: Deliberation, Motivation and the Nature of Value*, Cambridge: Cambridge University Press.
HILLS, A., 2009, "Moral Testimony and Moral Epistemology," *Ethics*, 120 (1): 94-127.
HILLS, A., 2016, "Understanding Why," *Noûs*, 50 (4): 661-688.
HOWARD, N. R., 2018, "Sentimentalism About Moral Understanding," *Ethical Theory and Moral Practice*, 21 (5): 1065-1078.
HUEMER, M., 2001, *Skepticism and the Veil of Perception*, Rowman& Littlefield.
HUEMER, M., 2005, *Ethical Intuitionism*, Basingstoke: Palgrave Macmillan.
HUME, D., 1739-40 [T], *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd rev. edn., P. H. Nidditch. Oxford: Clarendon Press, 1978.
HUME, D., 1751, *An Enquiry Concerning the Principles of Morals*. Ed. J. B. Schneewind. Indianapolis, Hackett, 1983.
HUTCHESON, F., 1725, *An Inquiry Into the Original of Our Ideas of Beauty and Virtue in Two Treatises*, W. Leidhold (ed.), Indianapolis, IN: Liberty Fund,

2004. [Hutcheson 1725 available online]
HUTCHESON, F., 1728, *An Essay on the Nature and Conduct of Passions, With Illustrations on the Moral Sense*, A. Garrett (ed.), Indianapolis, IN: Liberty Fund, 2002. [Hutcheson 1728 available online]
JACOBSON, D., 2012, "Moral Dumbfounding and Moral Stupefaction," *Oxford Studies in Normative Ethics,* 2: 289-316.
JAMES, A., 2007, "Constructivism About Practical Reasons," *Philosophy and Phenomenological Research*, 74(2): 302-325.
JOHNSTON, M., 1989, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, 63(Supplement): 139-174.
JOYCE, R., 2013, "Error Theory," in H. LaFollette (ed.), *The International Encyclopedia of Ethics*, Hoboken, NJ: John Wiley & Sons.
KAHANE, G., 2012, "On the Wrong Track: Process and Content in Moral Psychology," *Mind and Language*, 27(5): 519-545.
KAHNEMAN, D. and G. KLEIN, 2009, "Conditions for Intuitive Expertise: A Failure to Disagree," *American Psychologist*, 64(6): 515-526.
KAUPPINEN, A., 2010, "What Makes a Sentiment Moral?" *Oxford Studies in Metaethics*5: 225-256.
KAUPPINEN, A., 2013, "A Humean Theory of Moral Intuition," *Canadian Journal of Philosophy*, 43(3): 360-381.
KAUPPINEN, A., 2014a, "Empathy, Emotion Regulation, and Moral Judgment," in *Empathy and Morality*, H. Maibom (ed.), Oxford: Oxford University Press, pp. 97-121.
KAUPPINEN, A., 2014b, "Fittingness and Idealization," *Ethics* 124(3): 572-588.
KAUPPINEN, A., 2015, "Intuition and Belief in Moral Motivation," in *Motivational Internalism*, G. Björnsson, C.
STRANDBERG, R. F. OLINDER, J. ERIKSSON, and F. BJÖRKLUND (eds.), Oxford: Oxford University Press, pp. 237-259.
KENNY, A., 1963, *Action, Emotion, and the Will*. London: Routledge and Kegan Paul.
KOPAJTIC, L., 2020, "Adam Smith's Sentimentalist Conception of Self-Control," *Adam Smith Review*, 12: 7-27.
KORSGAARD, C., 1986, "SkepticismAbout Practical Reason," *Journal of Philosophy*, 83(1): 5-25.
KRIEGEL, U., 2012, "Moral Motivation, Moral Phenomenology, and the Alief/Belief Distinction," *Australasian Journal of Philosophy*, 90(3): 469-486.

KRIEGEL, U., 2014, "Towards a New Feeling Theory of Emotion," *European Journal of Philosophy* 22(3): 420-442.
KOONS, J.R., 2003, "Why Response-Dependence Theories of Morality are False," *Ethical Theory and MoralPractice*, 6(3): 275-294.
LOCKE, J., 1694 [1975], *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Clarendon Press (original work, 2nd ed., first published 1694).
LENMAN, J., 1999, "The Externalist and the Amoralist," *Philosophia*, 27(3-4): 441-457.
LEWIS, D. K., 1989, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, 63(Supplement): 113-137.
LONG, A. A. and D. N. SEDLEY, 1987, *The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary*, Cambridge: Cambridge University Press.
LYONS, W., 1980, *Emotion*, Cambridge: Cambridge University Press.
MACKIE, J., 1980, *Hume's Moral Theory*, London: Routledge and Kegan Paul.
MAIBOM, H. L., 2005, "Moral Unreason: The Case of Psychopathy," *Mind and Language*, 20 (2): 237-257.
MARKIE, P., 2005, "The Mystery of Direct Perceptual Justification," *Philosophical Studies*, 126 (3): 347-373.
MAY, J., 2018, *Regard for Reason in the Moral Mind*, New York: Oxford University Press.
MCDOWELL, J., 1978, "Are Moral Requirements Hypothetical Imperatives?",*Proceedings of the Aristotelian Society*, 52(Supplement): 13-29. Reprinted in McDowell 1998: 95-111.
MCDOWELL, J., 1979, "Virtue and Reason," *The Monist*, 62: 331-350.Reprinted in McDowell 1998: 50-73.
MCDOWELL, J., 1985, "Values and Secondary Qualities," in *Morality and Objectivity*, T. Honderich (ed.), London: Routledge and Kegan Paul, pp. 110-129; reprinted in McDowell 1998: 131-150.
MCDOWELL, J., 1987, "Projection and Truth in Ethics,", pamphlet, Department of Philosophy, University of Kansas; reprinted in McDowell 1998: 151-166.
MCDOWELL, J., 1995, "Might There Be External Reasons?" in *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, J.E.J. Altham and R. Harrison (eds.), Cambridge: CambridgeUniversity Press, pp. 387-398; reprinted in McDowell 1998: 95-111.
MCDOWELL, J., 1998, *Mind, Value, and Reality*, Cambridge, MA: Harvard University Press.
MCNAUGHTON, D., 1988, *Moral Vision*, Oxford: Blackwell.

MILONA, M. and Naar, H., 2020, “Sentimental Perceptualism and the Challenge from Cognitive Bases,” *Philosophical Studies*, 177(10): 3071-3096.
MITCHELL, J., 2017, “The Epistemology of Emotional Experience,” *Dialectica*71(1):57-84.
MITCHELL, J., forthcoming, “The Bodily Attitudinal Theory of Emotion,” *Philosophical Studies*.
MOLL, J., R. de Oliveira-Souza, and P.J. Eslinger, 2003, “Morals and the Human Brain: A Working Model,” *Neuroreport*, 14(3): 299-305.
MOORE, G.E., 1903, *Principia Ethica*, Cambridge: Cambridge University Press.
MOORE, G.E., 1912, *Ethics*, London: Williams and Norgate.
MURDOCH, I. 1970. *Sovereignty of the Good*. London: Routledge and Kegan Paul.
NICHOLS, S., 2004, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford: Oxford University Press.
NICHOLS, S., 2005, “Innateness and Moral Psychology,” in *The Innate Mind: Structure and Content*, P. Carruthers, S.
LAURENCE, and S. STICH (eds.), New York: Oxford University Press, pp. 353-430.
ODDIE, G., 2005, *Value, Reality, and Desire*, Oxford University Press on Demand.
ODDIE, G., 2016, “Fitting Attitudes, Finkish Goods, and Value Appearances,” *Oxford Studies in Metaethics*, 11: 74-101.
PARFIT, D., 2011, *On What Matters*, Oxford: Oxford University Press.
PELSER, A. C., 2014, “Emotion, Evaluative Perception, and Epistemic Justification,” in Roeser and Todd (eds.) 2014: 107-123.
PETTIT, P., 1991, “Realism and Response-Dependence,” *Mind*, 100: 587-626.
PRINZ, J., 2004, *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
PRINZ, J., 2006, “The Emotional Basis of Moral Judgments,” *Philosophical Explorations*, 9(1): 29-43.
PRINZ, J., 2007, *The Emotional Construction of Morals*, New York: Oxford University Press.
PRINZ, J., 2011, “Against Empathy,” *Southern Journal of Philosophy*, 49: 214-233.
RABINOWICZ, W. and T. RØNNOW-RASMUSSEN, 2004, “The Strike of the Demon: On Fitting Pro-Attitudes and Value,” *Ethics*, 114(3): 391-423.
RADCLIFFE, E., 2006, “Moral Internalism and Moral Cognitivism in Hume’s Metaethics,” *Synthese*, 152(3): 353-370.
RAILTON, P., 2003, “Red, Bitter, Good,” in *Facts, Values, and Norms*. Cambridge, Cambridge University Press, 131-147.
RAILTON, P., 2014, “The Affective Dog and Its Rational Tale: Intuition and Attunement,” *Ethics*, 124 (4): 813-859.
RIDGE, M., 2006, “Ecumenical Expressivism: Finessing Frege,” *Ethics*, 116(2): 302-336.

RIDGE, M., 2014, *Impassioned Belief*. Oxford: Oxford University Press.
ROBERTS, R. C., 2003, *Emotions: An Essay in Aid of Moral Psychology*, Cambridge University Press.
ROESER, S., 2011, *Moral Emotions and Intuitions*, Basingstoke: Palgrave Macmillan.
ROESER, S. and C. TODD, 2014, *Emotion and Value*. Oxford: Oxford University Press.
SALMELA, M., 2011, "Can Emotion Be Modelled on Perception?" *dialectica*, 65(1): 1-29.
SAYRE-MCCORD, G., 1994, "On Why Hume's General Point of View Isn't Ideal - And Shouldn't Be," *Social Philosophy and Policy*, 11(1): 202-228.
SCANLON, T. M. 1998, *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
SCHROEDER, M., 2009, "Hybrid Expressivism: Virtues and Vices," *Ethics*, 119: 257-309.
SCHROEDER, M., 2010, *Non-Cognitivism in Ethics*, London: Routledge.
SCHROEDER, M., 2013, "Tempered Expressivism," *Oxford Studies in Metaethics*, 8, 283-314.
SHAFER-LANDAU, R., 2003, *Moral Realism: A Defence*, Oxford: Oxford University Press.
SHAFTESBURY, Third Earl of (Anthony Ashley Cooper), 1699-1714, *An Inquiry Into Virtue and Merit*, in D. Den Uyl (ed.), *Characteristicks of Men, Manners, Opinions, Times, Vol.2*, Indianapolis, IN: Liberty Fund, 2001, pp. 1-100.
SINHABABU, N., 2017, *Humean Nature: How Desire Explains Action, Thought, and Feeling*, Oxford University Press.
SINNOTT-ARMSTRONG, W. (ed.), 2008, *Moral Psychology. Vols. 1-3*, Cambridge, MA: MIT Press.
SLOTE, M., 2010, *Moral Sentimentalism*, Oxford: Oxford University Press.
SMITH, A., 1759-1790 [TMS], *The Theory of Moral Sentiments*, K. Haakonssen (ed.), Cambridge: Cambridge University Press 2002.
SMITH, M., 1987, "The Humean Theory of Motivation," *Mind*, 96: 36-61.
SMITH, M., 1994, *The Moral Problem*, Oxford: Blackwell.
SMITH, M., 1997, "In Defence of *The Moral Problem*," *Ethics*, 108: 84-119.
SOBEL, D. and D. COPP, 2001, "Against Direction of Fit Accounts of Belief and Desire," *Analysis*, 61 (269): 44-53.
SOBEL, D., 2009, "Subjectivism and Idealization," *Ethics*, 119: 336-352.
SOBER, E. and D.S. WILSON, 2008, *Unto Others.The Evolution and Psychology of Unselfish Behaviour*, Cambridge, MA: Harvard University Press.
SOSA, E., 1999, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13: 141-153.

STAMPE, D. W., 1987, "The Authority of Desire," *The Philosophical Review*, 96 (3): 335-381.

Stevenson, C., 1937, "The Emotive Meaning of Ethical Terms," *Mind*, 46: 14-31.

STEVENSON, C., 1944, *Ethics and Language*, New Haven and London: Yale University Press.

TAPPOLET, C., 2000, *EmotionsetValeurs*, Paris: PressesUniversitaires de France.

TAPPOLET, C., 2011, "Values and Emotions: Neo-Sentimentalism's Prospects," in Bagnoli (ed.) 2011: 117-134.

TAPPOLET, C., 2016, *Emotions, Values, and Agency*, Oxford University Press.

TOPPINEN, T., 2013, "Believing in Expressivism," *Oxford Studies in Metaethics*, 8, 252-282.

TUCKER, C., (ed.), 2013, *Seemings and Justification*, Oxford: Oxford University Press.

VANROOJEN, M., "Moral Intuitionism, Experiments and Skeptical Arguments," in A. Booth and D. Rowbottom (ed.), *Intuitions*, Oxford University Press, pp. 148-164.

WAY, J., 2012, "Transmission and the Wrong Kind of Reason," *Ethics*, 22(3): 489-515.

WESTERMARCK, E., 1906, *The Origin and Development of Moral Ideas, Vol. 1*, London: MacMillan & Co.

WHEATLEY, T. and J. HAIDT, 2005, "Hypnotically Induced Disgust Makes Moral Judgments More Severe," *Psychological Science*, 16: 780-784.

WIGGINS, D., 1987, "A Sensible Subjectivism?" in *Needs, Values, and Truth*. Oxford: Blackwell, pp. 185-214.

WILLIAMS, B., 1979, "Internal and External Reasons," reprinted in *Moral Luck*, Cambridge: Cambridge University Press, 1981, pp. 101-113.

WISNEWSKI, J. J., 2015, "The Case for Moral Perception," *Phenomenology and the Cognitive Sciences*, 14 (1): 129-148.

WOOD, A., 2011, "Humanity as an End in Itself," in Parfit 2011: vol. 2, pp. 58-82.

WRIGHT, C., 1988, "Moral Values, Projections, and Secondary Qualities," *Proceedings of the Aristotelian Society*, 62(Supplement): 1-26.

WRIGHT, C., 1992, *Truth and Objectivity*, Cambridge, MA: Harvard University Press.

YOUNG, L. and L. TSOI, 2013, "When Mental States Matter, When They Don't, and What That Means for Morality," *Social and Personality Psychology Compass*, 7 (8): 585-604.

ZAGZEBSKI, L., 2003, "Emotion and Moral Judgment," *Philosophy and Phenomenological Research*, 66(1): 104-124.

(V) Psicologia Evolutiva*

Autoria: Stephen M. Downes
Tradução: Sagid Salles
Revisão: Roseline Crippa

A psicologia evolutiva é uma das muitas abordagens biologicamente informadas do estudo do comportamento humano. Juntamente com os psicólogos cognitivos, os psicólogos evolutivos propõem que muito de nosso comportamento, se não todo ele, pode ser explicado apelando-se a mecanismos psicológicos internos. O que distingue os psicólogos evolutivos de muitos psicólogos cognitivos é a proposta de que os mecanismos internos relevantes são adaptações - produtos da seleção natural - que ajudaram nossos ancestrais a lidar com o mundo, a sobreviver e a reproduzir. Para compreender as teses centrais da psicologia evolutiva, precisamos compreender alguns conceitos-chave em psicologia evolutiva, psicologia cognitiva,

---

* DOWNES, S. M. Evolutionary Psychology. *In*: ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy.** Spring Edition. Stanford, CA: The Metaphysics Research Lab. Disponível em: https://plato.stanford.edu/archives/spr2021/entries/evolutionary-psychology/. Acesso em: 18 jul. 2022.

The following is the translation of the entry on Evolutionary Psychology by Stephen M. Downes, 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/spr2021/entries/evolutionary-psychology/. 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 a https://plato.stanford.edu/entries/evolutionary-psychology/. 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.

filosofia da ciência e filosofia da mente. Filósofos estão interessados na psicologia evolutiva por diversas razões. Para filósofos da ciência - sobretudo filósofos da biologia - a psicologia evolutiva proporciona um alvo de críticas. Embora haja, entre os filósofos da biologia, um amplo consenso de que a psicologia evolutiva é um empreendimento profundamente falho, isto não implica que estes filósofos rejeitam completamente a relevância da teoria evolutiva para a psicologia humana. Para filósofos da mente e para a ciência cognitiva, a psicologia evolutiva tem sido uma fonte de hipóteses empíricas sobre a arquitetura cognitiva e componentes específicos desta arquitetura. Entretanto, alguns filósofos da mente também são críticos da psicologia evolutiva, embora suas críticas não sejam tão abrangentes quanto aquelas apresentadas pelos filósofos da biologia. Filósofos interessados em psicologia moral também recorrem à psicologia evolutiva, tanto como uma fonte de hipóteses empíricas quanto como um alvo de críticas.

A seguir, eu explico, brevemente, as relações da psicologia evolutiva com outros trabalhos sobre a biologia do comportamento humano e as ciências cognitivas. Após isto, introduzo os principais conceitos teóricos da tradição de pesquisa. Na seção seguinte, abordo discussões sobre a psicologia evolutiva na filosofia da mente, focando-me especificamente no debate sobre a tese da modularidade massiva. Passo a examinar algumas das críticas à psicologia evolutiva apresentadas pelos filósofos da biologia, e considero algumas respostas a estas críticas. Eu então apresento algumas das contribuições da psicologia evolutiva para a psicologia moral e a natureza humana, e, finalmente, discuto brevemente o alcance e impacto da psicologia evolutiva.

1. Psicologia Evolutiva: Uma tradição de pesquisa entre as várias abordagens biológicas para explicar o comportamento humano

O foco deste verbete é a abordagem específica à psicologia evolutiva convencionalmente chamada “Psicologia Evolutiva”, com iniciais maiúsculas. A ideia de convencionar este nome é de David Buller (2000, 2005). Ele introduz a convenção para distinguir uma tradição particular de pesquisa (LAUDAN, 1977) de outras abordagens à biologia do comportamento humano.[54] Embora esta tradição de

[54] Buller se refere à “psicologia evolutiva” como um “paradigma”. Nós adotamos a terminologia de “tradição de pesquisa”, de Laudan. Tradições de pesquisa têm uma estrutura mais fluída do que paradigmas, e Laudan permite o compartilhamento de recursos teóricos entre tradições de pesquisa.

pesquisa seja o foco aqui, letras minúsculas são usadas ao longo de todo o texto, pois nenhum outro tipo de psicologia evolutiva é discutido. A psicologia evolutiva se baseia em preceitos teóricos específicos (apresentados na seção 2, abaixo), e nem todos estes preceitos são compartilhados por outras pessoas que trabalham com a biologia do comportamento humano (LALAND; BROWN, 2002; BROWN *et al*., 2011). Por exemplo, ecologistas comportamentais humanos apresentam e defendem hipóteses de explicação acerca do comportamento humano que não apelam a mecanismos psicológicos (*vide* HAWKES, 1990; HRDY, 1999). Ecologistas comportamentais também acreditam que muito do comportamento humano pode ser explicado apelando-se à evolução, enquanto rejeitam a ideia, sustentada pelos psicólogos evolutivos, de que um período de nossa história evolutiva é a origem de todas as nossas adaptações psicológicas importantes (IRONS,1998). Psicobiólogos do desenvolvimento adotam ainda outra abordagem: eles são antiadaptacionistas. (MICHEL; MOORE, 1995; mas, para exemplos de trabalhos desenvolvimentistas com um viés adaptacionista, veja BATESON; MARTIN, 1999; BJORKLUND; BLASI, HERNANDEZ, 2005). Estes teóricos acreditam que muito de nosso comportamento pode ser explicado sem o apelo a um conjunto de adaptações psicológicas específicas para o comportamento em questão. Ao invés disto, eles enfatizam o papel do desenvolvimento na produção de vários traços comportamentais humanos. Daqui em diante, "psicologia evolutiva" se refere a uma tradição específica de pesquisa, uma dentre as muitas abordagens biológicas ao estudo do comportamento humano.

Paul Griffiths argumenta que a psicologia evolutiva tem uma dívida teórica com a sociobiologia e a etologia (GRIFFITHS, 2006, 2008). Psicólogos evolutivos reconhecem sua dívida com a sociobiologia, mas ressaltam que acrescentaram uma dimensão a ela: os mecanismos psicológicos. Comportamentos humanos não são um produto direto da seleção natural, mas sim o produto de mecanismos psicológicos que foram selecionados. A relação com a etologia aqui é que, na década de 1950, etologistas apresentaram instintos ou impulsos subjacentes a nosso comportamento;[55] os mecanismos psicológicos da psicologia evolutiva são os correlatos aos instintos ou impulsos. A psicologia evolutiva também está relacionada com a psicologia cognitiva e as ciências cognitivas. Os mecanismos psicológicos aos quais elas recorrem são computacionais, às vezes referidos como "algoritmos

[55] Desde então, a etologia se afastou dessa linha de pensamento (ERESHEFSKY, 2007).

darwinianos" ou "módulos computacionais". Este cognitivismo explícito separa a psicologia evolutiva da neuroendocrinologia comportamental e de muitos trabalhos em neurociência. Embora nestas áreas mecanismos internos sejam propostos nas explicações do comportamento humano, tais mecanismos não são interpretados em termos computacionais. Frequentemente recorre-se à bem conhecida distinção tripartida de David Marr (*vide* MARR, 1983) para distinguir entre os níveis nos quais pesquisadores das ciências cognitivas e neurociências concentram sua atenção. Muitos neurocientistas e neuroendocrinologistas comportamentais trabalham no nível da implementação, enquanto psicólogos cognitivos trabalham no nível das computações implementadas no nível neurobiológico (*vide* GRIFFITHS, 2006).

Às vezes, psicólogos evolutivos apresentam sua abordagem como sendo potencialmente unificadora, ou como fornecendo uma base para todos os outros trabalhos que pretendem explicar o comportamento humano (*vide* TOOBY; COSMIDES, 1992). Esta posição tem sido recebida com forte ceticismo por muitos cientistas sociais, que veem um papel para uma infinidade de tipos de explicação do comportamento humano, algumas das quais irredutíveis a explicações biológicas de qualquer tipo. Esta discussão se conecta a questões acerca do reducionismo nas ciências sociais. (Little, de 1991, tem uma boa introdução a estas questões). Também há razões para acreditar que a psicologia evolutiva nem unifica nem fornece bases para áreas que são suas vizinhas próximas, como a ecologia comportamental ou a psicobiologia do desenvolvimento. (*vide* DOWNES, 2005). Em outros trabalhos, psicólogos evolutivos apresentam sua abordagem como sendo consistente ou compatível com abordagens vizinhas, como a ecologia comportamental e a psicobiologia do desenvolvimento. (*vide* BUSS, 2005). A verdade desta posição depende de um exame cuidadoso dos preceitos teóricos da psicologia evolutiva e das áreas vizinhas. Voltamo-nos agora para os preceitos teóricos da psicologia evolutiva, e revisitamos esta discussão na seção 4, abaixo.

2. Teoria e métodos da psicologia evolutiva

Psicólogos evolutivos influentes, como Leda Cosmides e John Tooby, fornecem a seguinte lista de preceitos teóricos da área (TOOBY; COSMIDES, 2005).

1 O cérebro é um computador projetado pela seleção natural para extrair informação do ambiente.

2 O comportamento humano individual é gerado por este computador evoluído como resposta à informação que ele extrai do ambiente. A compreensão do comportamento requer a articulação dos programas cognitivos que geram o comportamento.
3 Os programas cognitivos do cérebro humano são adaptações. Eles existem porque produziram, em nossos ancestrais, comportamentos que os permitiram sobreviver e reproduzir.
4 Os programas cognitivos do cérebro humano podem não ser atualmente adaptativos; eles eram adaptativos em ambientes ancestrais.
5 A seleção natural faz com que o cérebro seja composto de muitos programas com propósitos específicos diferentes, e não o domínio de uma arquitetura geral.
6 A descrição da arquitetura computacional evoluída de nossos cérebros "permite uma compreensão sistemática dos fenômenos culturais e sociais" (TOOBY; COSMIDES, 2005, p. 16-18).

O preceito 1 ressalta o cognitivismo com o qual psicólogos evolutivos estão comprometidos. Em combinação com 2, o preceito 1 direciona nossa atenção, enquanto pesquisadores, não para as partes do cérebro, mas para os programas executados por ele. Estes programas - os mecanismos psicológicos - é que são produtos da seleção natural. Embora sejam produtos da seleção natural, e consequentemente adaptações, estes programas não precisam ser atualmente adaptativos. Nosso comportamento pode ser produzido por mecanismos psicológicos subjacentes que surgiram em resposta a circunstâncias particulares do ambiente de nossos ancestrais. O preceito 5 apresenta o que é frequentemente chamado "tese da modularidade massiva" (*vide* SAMUELS, 1998, 2000). Há muitas coisas contidas neste preceito, e examinaremos esta tese com algum detalhe abaixo, na seção 3. Em resumo, psicólogos evolutivos sustentam que existe uma analogia entre órgãos e módulos ou mecanismos psicológicos. Órgãos realizam bem funções específicas e são o produto da seleção natural. Não existem órgãos com propósitos gerais, corações bombeiam o sangue e fígados desintoxicam o sangue. Ele se aplica aos mecanismos psicológicos; eles surgem como respostas a contingências específicas do ambiente, e são selecionados na medida em que contribuem para

a sobrevivência e reprodução do organismo. Assim como não existem órgãos com propósitos gerais, não existem mecanismos psicológicos com propósitos gerais. Por fim, o preceito 6 introduz a perspectiva reducionista ou fundacional da psicologia evolutiva, que foi discutida acima.

Existem inúmeros exemplos de tipos de mecanismos que, de acordo com pesquisas guiadas por estes preceitos teóricos, hipoteticamente subjazem a nosso comportamento: o módulo de detecção de trapaceiros; o módulo de leitura da mente; o módulo de detecção da relação cintura-quadril; o módulo de medo de cobras; e assim por diante. Um olhar mais atento ao módulo de detecção da relação cintura-quadril ilustra os preceitos teóricos acima em ação. Devendra Singh (SINGH, 1993, SINGH; LUIS, 1995) apresenta o modulo de detecção da relação cintura-quadril como pertencendo ao conjunto de módulos que subjazem à escolha de parceiros em humanos. Este é um mecanismo psicológico particularmente masculino. Homens detectam variações na relação cintura-quadril das mulheres. A preferência dos primeiros é por mulheres com uma relação cintura-quadril mais próxima a .7. Singh afirma que a detecção e o conjunto de preferências são adaptações para escolher parceiros férteis. Deste modo, nosso comportamento ao selecionar parceiros é parcialmente explicado pelos mecanismos psicológicos subjacentes a nossas preferências com respeito à relação cintura-quadril. Mecanismos estes que foram selecionados em ambientes humanos passados.

O que importa notar acerca da pesquisa guiada pelos preceitos teóricos acima é que todo comportamento é melhor explicado em termos de mecanismos psicológicos subjacentes, que são adaptações para resolver um conjunto particular de problemas com os quais nossos ancestrais humanos tiveram de lidar em algum momento. Além disto, psicólogos evolutivos enfatizam que os mecanismos nos quais eles se concentram são universalmente distribuídos em humanos, e não são suscetíveis a muita variação, se é que são suscetíveis a alguma variação. Eles sustentam que os mecanismos, embora sejam um produto da adaptação, não estão mais sujeitos à seleção (TOOBY; COSMIDES, 2005, p. 39-40). A introdução acessível e abrangente de Clark Barrett (2015) à psicologia evolutiva defende esta ênfase em mecanismos evoluídos como o principal foco das pesquisas em psicologia evolutiva. Barrett ainda expande o escopo da psicologia evolutiva e percebe a adição de métodos de pesquisa desenvolvidos desde que Cosmides e Tooby primeiramente estabeleceram os parâmetros para a pesquisa nesta área. Algumas das propostas de Barrett são discutidas nas seções 6 e 7, abaixo. Todd Shackleford e Viviana Weekes-Shackleford (2017) acabaram de completar um enorme compêndio de

trabalhos nas ciências psicológicas de base evolutiva. Um vasto leque de métodos diferentes de pesquisa é apresentado e defendido neste volume, e existem diversos verbetes comparando os méritos de abordagens alternativas à psicologia evolutiva.

Os métodos para testar hipóteses em psicologia evolutiva vêm principalmente da psicologia. No trabalho de Singh, por exemplo, sujeitos do sexo masculino são apresentados a figuras de mulheres com diferentes relações cintura-quadril, e são solicitados a fornecer classificações de suas preferências. Em seu trabalho corroborando vários mecanismos hipotéticos de seleção de parceiros, Buss (1990) realizou experimentos similares com sujeitos, solicitando suas respostas a várias questões sobre características dos parceiros desejados. Buss, Singh e outros psicólogos evolutivos enfatizam a validade transcultural de seus resultados, alegando consistência nas respostas em uma ampla variedade de populações humanas. Confira Yu e Shepard (1998) e Gray *et al*. (2003) para diferentes tipos de resultados que conflitam com os de Singh. Na maior parte das vezes, métodos experimentais psicológicos padrão são usados para testar hipóteses em psicologia evolutiva. Isto levantou dúvidas acerca da medida em que o componente evolutivo das hipóteses da psicologia evolutiva está sendo testado. (*vide* SHAPIRO; EPSTEIN, 1998; LLOYD, 1999; LLOYD; FELDMAN, 2002). Um perfil de resposta pode prevalecer em uma ampla variedade de populações de sujeitos, mas isto nada diz sobre se o perfil de resposta é ou não é um mecanismo psicológico que surgiu de um regime seletivo particular.

3. A hipótese da modularidade massiva

Alegações de que a mente tem uma arquitetura modular, ou mesmo massivamente modular, são difundidas na ciência cognitiva (*vide* HIRSHFIELD; GELMAN, 1994). A tese da modularidade massiva é, antes de mais nada, uma tese sobre a arquitetura cognitiva. Tal como defendida por psicólogos evolutivos, esta tese também diz respeito à origem de nossa arquitetura cognitiva: a arquitetura massivamente modular é o resultado da seleção natural agindo para produzir cada um dos muitos módulos (*vide* BARRETT; KURZBAN, 2006; BARRETT, 2012). Nossa arquitetura cognitiva é composta de dispositivos computacionais inatos que são adaptações (*vide* SAMUELS, 1998, 1999a, 1999b, 2000). Esta arquitetura massivamente modular responde por todo nosso comportamento sofisticado. Nossa navegação bem-sucedida no mundo resulta da ação de um ou mais de nossos muitos módulos.

Jerry Fodor foi o primeiro a montar uma defesa filosófica embasada da modularidade como uma teoria da arquitetura cognitiva (FODOR, 1983). Sua tese da modularidade distingue-se da tese da modularidade massiva de várias maneiras importantes. Fodor argumentou que nossos "sistemas de input" são modulares - por exemplo, componentes de nosso sistema visual, nosso sistema de detecção de fala e assim por diante. Estas partes de nossas mentes são processadores especializados de informação, cuja composição interna é inacessível a outros processadores relacionados. Os sistemas de detecção modular fornecem output a um sistema central, que é uma espécie de mecanismo de inferência. Na perspectiva de Fodor, o sistema central não é modular. Fodor apresenta um grande número de argumentos contra a possibilidade de sistemas modulares centrais. Por exemplo, ele argumenta que sistemas centrais, na medida em que se ocupam de algo como a confirmação científica, são quineanos, no sentido de que "o grau de confirmação atribuído a qualquer hipótese dada é sensível a propriedades do sistema de crenças como um todo" (FODOR, 1983, p. 107). De seu exame das características dos sistemas centrais, Fodor extrai uma conclusão desoladora acerca do status da ciência cognitiva: a ciência cognitiva é impossível. Deste modo, na perspectiva de Fodor a mente é parcialmente modular, e a parte modular fornece alguma matéria para a ciência cognitiva.

A tese da modularidade massiva, distinta da de Fodor, recebe uma defesa filosófica embasada de Peter Carruthers (*vide* CARRUTHERS, 2006). Carruthers está bem ciente de que Fodor não acredita que sistemas centrais podem ser modulares (*vide* FODOR, 2000). Mas ele apresenta argumentos, tanto de psicólogos evolutivos como de outros, que apoiam a tese da modularidade aplicada a toda a mente. Talvez uma das razões de haver tanto interesse filosófico na psicologia evolutiva seja que discussões sobre o status da tese da modularidade massiva são altamente teóricos.[56] Tanto psicólogos evolutivos como filósofos apresentam e consideram argumentos a favor e contra a tese em questão, ao invés de simplesmente esperar até que os resultados empíricos apareçam. Richard Samuels (1998) especula

---

[56] Existe semelhança entre o caso a favor da modularidade massiva e o caso a favor da hipótese da linguagem do pensamento. Ambos se baseiam em argumentação. Austin Booth (2004) argumentou que, o caso a favor da última, tem a aparência de um argumento transcendental.

que argumentos são mais confiáveis do que dados empíricos, porque as várias teses concorrentes da modularidade dos sistemas centrais são difíceis de separar empiricamente. Carruthers é um exemplo desta abordagem, pois se apoia fortemente em argumentos a favor da modularidade massiva, frequentemente à custa de resultados empíricos específicos que contam a favor desta tese.

Existem muitos argumentos a favor da tese da modularidade massiva. Alguns são baseados em considerações sobre como a evolução deve ter agido; alguns são baseados em considerações sobre a natureza da computação e alguns são versões do argumento da pobreza de estímulo, apresentado primeiramente por Chomsky a favor da existência de uma gramática universal inata. Confira Cowie (1999) para uma boa apresentação da estrutura dos argumentos da pobreza de estímulo. Uma infinidade de versões de cada um destes argumentos aparece na literatura, e muitos argumentos a favor da modularidade massiva mesclam e combinam componentes de cada uma das principais linhas de argumentação. Avaliaremos uma versão de cada tipo de argumento aqui.

Carruthers apresenta um claro esboço do primeiro tipo de argumento: o "argumento biológico a favor da modularidade massiva". "(1) Sistemas biológicos são projetados, construídos de forma gradual. (2) Quando são complexos, tais sistemas precisam ter uma organização massivamente modular. (3) A mente humana é um sistema biológico complexo. (4) Assim, a organização da mente humana será massivamente modular" (CARRUTHERS, 2006, p. 25). Um exemplo deste argumento consiste em apelar à decomposição funcional de organismos em órgãos "projetados" para tarefas específicas, por exemplo, corações, fígados e rins. Cada um destes órgãos surge como o resultado da seleção natural, e, agindo em conjunto, eles contribuem para a aptidão do organismo. A decomposição funcional é impulsionada pela resposta a estímulos ambientais específicos. Ao invés de a seleção natural agir para produzir órgãos com propósitos gerais, cada desafio ambiental específico é encarado por um mecanismo separado. Todas as versões deste argumento são argumentos por analogia, que se apoiam na premissa intermediária chave de que mentes são um tipo de sistema biológico sobre o qual a seleção natural atua.

O Segundo tipo de argumento não apela a quaisquer considerações biológicas (embora muitos psicólogos evolutivos deem um retoque biológico a estes argumentos). Chame este argumento de "argumento computacional". Ele se desenvolve do seguinte modo: mentes são dispositivos computacionais voltados para a solução de problemas; existem tipos específicos de soluções para tipos específicos de problemas; deste modo, para que mentes sejam dispositivos gerais

(e bem-sucedidos) de solução de problemas, elas têm de consistir em coleções de dispositivos voltados para a solução de problemas específicos, ou seja, de muitos módulos computacionais. Este tipo de argumento é estruturalmente similar ao argumento biológico (como aponta Carruthers). A ideia central é que não existe sentido na ideia de um solucionador geral de problemas, e nenhum progresso pode ser feito na ciência cognitiva sem decompor os problemas em suas partes.

O terceiro tipo de argumento envolve uma generalização do argumento da pobreza de estímulo, de Chomsky, a favor de uma gramática universal. Muitos psicólogos evolutivos (*vide* TOOBY; COSMIDES, 1992) apelam à ideia de que não há nem tempo nem informações disponíveis suficientes para qualquer humano aprender do zero a resolver com sucesso todos os problemas com os quais nos deparamos no mundo. Esta consideração preliminar apoia a conclusão de que os mecanismos subjacentes que usamos para resolver os problemas relevantes são inatos (para psicólogos evolutivos, "inato" normalmente é intercambiável com "produto da seleção natural"[57]). Se aplicarmos este argumento a toda gama de conjuntos de problemas que humanos enfrentam e resolvem, obtemos um enorme conjunto de mecanismos inatos que servem às nossas habilidades de resolver problemas. Este é outro modo de dizer que temos uma mente massivamente modular.

Existem inúmeras respostas às muitas versões de cada um destes tipos de argumentos, e muitos assumem abertamente a tese da modularidade massiva sem considerar um argumento específico a seu favor. Vou adiar a consideração de respostas ao primeiro tipo de argumento até a seção 4, abaixo. A seção 4 se concentra em questões sobre a natureza da evolução e da seleção natural - temas da filosofia da biologia.

O segundo tipo de argumento é uma face de um debate perene na filosofia da ciência cognitiva. De acordo com Fodor (2000, p. 68), este argumento se apoia na suposição injustificada de que não existe um critério de sucesso cognitivo que é independente de domínios, o que ele pensa requerer um argumento não apresentado pelos psicólogos evolutivos. Em resposta aos psicólogos evolutivos, Samuels (*vide* SAMUELS, 1998) alega que argumentos deste tipo não discriminam suficientemente entre uma conclusão sobre mecanismos de processamento específicos a domínios e conhecimento ou informação específica a domínios. Samuels articula o que ele

---

[57] Psicólogos evolutivos também propõem outras explicações do inato.

chama de “modelo da cognição como uma biblioteca”, no qual existe informação ou conhecimento específico a domínios, mas processamento geral a domínios. Embora o modelo da cognição como uma biblioteca não seja massivamente modular no sentido relevante, ele é apoiado pelo segundo tipo de argumento. De acordo com Samuels, psicólogos evolutivos precisam de algo além deste tipo de argumento para justificar seu tipo específico de conclusão sobre a modularidade massiva. Buller (2005) introduz preocupações adicionais para este tipo de argumento ao atacar a suposição de que não pode haver algo como um mecanismo para a solução de problemas que seja geral a domínios.

Buller receia que, em sua tentativa de apoiar esta tese, psicólogos evolutivos falham em caracterizar adequadamente um solucionador de problemas geral a domínios. Por exemplo, eles falham em distinguir entre um solucionador de problemas geral a domínios e um solucionador de problemas específico a domínios que é indevidamente generalizado. Ele oferece o exemplo da aprendizagem social como um mecanismo geral a domínios que produziria soluções para problemas que são específicas a domínios. Ele usa uma boa analogia biológica para enfatizar este ponto: o sistema imunológico é um sistema geral a domínios, na medida em que permite que o corpo responda a uma ampla variedade de patógenos. Embora seja verdade que o sistema imunológico produz, na forma de anticorpos específicos, respostas aos patógenos que são específicas a domínios, os anticorpos são produzidos por um sistema geral a domínios. Estes e muitos outros críticos concluem que os argumentos do segundo tipo não apoiam adequadamente a tese da modularidade massiva.

Fodor (2000) e Kim Sterelny (2003) fornecem respostas diferentes aos argumentos do terceiro tipo. A resposta de Fodor é que os argumentos da pobreza de estímulo apoiam conclusões sobre o inato, mas não sobre a modularidade, e, assim, eles não podem ser usados em apoio à tese da modularidade massiva. Ele argumenta que a especificidade e o encapsulamento de um mecanismo a domínios são coisas muito claramente separadas de seu caráter inato. Isto abre espaço para “mecanismos de aprendizagem perfeitamente gerais” que são inatos, “mecanismos totalmente encapsulados” que são específicos a um estímulo único, e tudo o que houver entre estas opções. Sterelny responde à generalização do terceiro tipo de argumento. Ele considera a linguagem a exceção, e não a regra, no sentido de que embora possa haver justificativa para explicar nossas habilidades linguísticas pela postulação de um módulo inato que seja específico a domínios, muitos de nossos outros comportamentos voltados para a solução de problemas podem ser explicados

sem a postulação de tais módulos (STERELNY, 2003, p. 200).[58] A resposta de Sterelny requer o apelo a explicações alternativas para nosso repertório de comportamentos. Por exemplo, ele explica a psicologia e a biologia de senso comum por meio do apelo a fatores ambientais, alguns dos quais produzidos por nossos antepassados, que nos permitem realizar tarefas cognitivas sofisticadas. Se pudermos explicar nosso sucesso em várias tarefas complexas voltadas para a solução de problemas sem apelar a módulos, então a tese da modularidade massiva fica prejudicada. Sterelny refina sua resposta à modularidade massiva ao adicionar mais detalhe a suas explicações de como muitos de nossos traços unicamente humanos podem ter evoluído (*vide* STERELNY, 2012). Para explicar a evolução de muitos traços humanos que muitos assumem requererem uma explicação em termos de modularidade massiva, como a formação de juízos morais por exemplo, ele introduz seu modelo do "aprendiz evoluído". Cecilia Heyes adota uma abordagem similar à de Sterelny ao atacar a modularidade massiva. Ao invés de apresentar argumentos contra a modularidade massiva, ela propõe explicações alternativas do desenvolvimento da psicologia de senso comum, que não se baseiam na tese da modularidade massiva (HEYES, 2014a, 2014b).

Heyes e Sterelny não apenas rejeitam a modularidade massiva, mas também têm baixa expectativa de que qualquer tese da modularidade dê frutos. Entretanto, existem muitos críticos da modularidade massiva que estão abertos à possibilidade de alguma modularidade da mente. Tais críticos da psicologia evolutiva não rejeitam a possibilidade de todo tipo de modularidade, eles rejeitam apenas a tese da modularidade massiva. Existe um debate considerável sobre o status da tese da modularidade massiva, e parte deste debate se concentra na caracterização dos módulos. Se módulos possuem todas as características primeiramente apresentadas por Fodor (1983), então pode ser correto que sistemas centrais não são modulares. Tanto Carruthers (2006) como Barrett e Kurzban (2006) apresentam caracterizações modificadas dos módulos, que eles argumentam servirem melhor à tese da modularidade massiva. Não existe acordo sobre uma caracterização dos módulos que seja viável para a psicologia evolutiva, mas existe acordo sobre a tese um tanto inofensiva de que "a linguagem da modularidade fornece uma base conceitual útil na qual podem ser traçados debates produtivos em torno dos sistemas cognitivos" (BARRETT; KURZBAN, 2006, p. 644).

---

[58] Buller (2005, p. 157-158) também apresenta uma versão deste argumento.

4. Filosofia da Biologia vs. Psicologia Evolutiva

Muitos filósofos têm criticado a psicologia evolutiva. A maioria destes críticos são filósofos da biologia, que argumentam que a tradição de pesquisa sofre de uma forma excessivamente entusiástica de adaptacionismo (GRIFFITHS, 1996; RICHARDSON, 1996; GRANTHAM; NICHOLS, 1999; LLOYD, 1999; RICHARDSON, 2007), de um reducionismo insustentável (DUPRE, 1999; DUPRE, 2001), de uma "má aposta empírica" sobre módulos (STERELNY, 1995; STERELNY; GRIFFITHS, 1999; STERELNY, 2003), de uma concepção apressada e pobre da aptidão (LLOYD, 1999; LLOYD; FELDMAN, 2002), e da maioria dos problemas acima e muito mais (BULLER; 2005) (*vide* DOWNES, 2005).[59] Todos estes filósofos compartilham uma ou outra versão da perspectiva de Buller: "Eu sou descaradamente entusiasmado com os esforços de aplicar a teoria evolutiva à psicologia humana" (2005, p. X).[60] Mas se, como citação de Buller indica, filósofos da biologia não são céticos com respeito à ideia fundamental por trás do projeto, então o que eles tanto criticam? O que está em jogo são perspectivas diferentes sobre a melhor forma de caracterizar a evolução, e, consequentemente, sobre como gerar e testar hipóteses evolutivas. Para psicólogos evolutivos, a contribuição mais interessante da teoria evolutiva é a explicação do aparente design na natureza, ou a explicação da produção de órgãos complexos pelo apelo à seleção natural. Psicólogos evolutivos geram hipóteses evolutivas primeiro encontrando um aparente design no mundo, por exemplo em nossa constituição psicológica, e em seguida apresentando um cenário seletivo que teria levado à produção do traço que exibe o aparente design. Dado que as hipóteses geradas pelos psicólogos evolutivos normalmente dizem respeito às nossas capacidades psicológicas, elas são testadas por métodos psicológicos padrão. Filósofos da biologia contestam os psicólogos evolutivos em ambos os pontos. Abaixo, apresento alguns poucos exemplos de críticas com respeito a cada um destes dois aspectos, e, em seguida, considero algumas respostas às críticas filosóficas à psicologia evolutiva.

---

[59] Este é um pequeno subconjunto da enorme literatura crítica sobre psicologia evolutiva.
[60] Richard Lewontin (1998) argumenta a favor de uma posição cética muito mais forte, que efetivamente afirma que a psicologia evolutiva é impossível.

Adaptação é o conceito biológico central à maioria dos debates sobre psicologia evolutiva. Todo trabalho teórico sobre psicologia evolutiva apresenta a tradição de pesquisa como primariamente focada nas adaptações psicológicas, e prossegue com uma explicação do que são adaptações (*vide* TOOBY; COSMIDES, 1992; BUSS *et al.*, 1998; SIMPSON; CAMPBELL, 2005; TOOBY; COSMIDES, 2005). Grande parte da crítica filosófica à psicologia evolutiva se dirige à sua abordagem da adaptação ou à sua forma de adaptacionismo. Analisemos rapidamente o fundamental da perspectiva da filosofia da biologia.

Eis como Elliott Sober define adaptação: "uma característica *c* é uma adaptação para realizar uma tarefa *t* em uma população se, e somente se, membros da população agora têm *c* porque, ancestralmente, houve seleção da posse de *c* e *c* conferiu uma vantagem em aptidão porque realizou a tarefa *t*" (SOBER, 2000, p. 85). Sober faz alguns esclarecimentos adicionais da noção de adaptação que são úteis. Primeiro, devemos distinguir entre um traço que é **adaptativo** e um traço que é uma **adaptação**. Inúmeros traços podem ser adaptativos sem que sejam adaptações. As patas dianteiras das tartarugas marinhas são úteis para cavar a areia e enterrar os ovos, mas não são adaptações para a construção de ninhos (SOBER, 2000, p. 85). Além disto, traços podem ser adaptações sem que sejam atualmente adaptativos para um dado organismo. Exemplos de tais traços são os órgãos vestigiais, como nosso apêndice e os olhos vestigiais em organismos que habitam cavernas (STERELNY; GRIFFITHS, p. 1999). Segundo, devemos distinguir entre adaptações ontogênicas e filogenéticas (SOBER, 2000, p. 86). As adaptações de interesse para biólogos evolutivos são as filogenéticas, que surgem ao longo do tempo evolutivo e impactam a aptidão do organismo. Embora adaptações ontogenéticas, incluindo qualquer comportamento que aprendemos ao longo de nossa vida, possam ser adaptativas na medida em que o organismo se beneficia delas, elas não são adaptações no sentido relevante. Por fim, "adaptação" e "função" são termos intimamente relacionados. De acordo com uma das perspectivas proeminentes das funções - a perspectiva etiológica das funções - "adaptação" e "função" são mais ou menos coextensivos; perguntar pela função de um organismo é perguntar por que ele existe atualmente. De acordo com a perspectiva de Cummins das funções, "adaptação" e "função" não são coextensivos, na medida em que perguntar qual a função de um órgão é perguntar o que ele faz (SOBER, 2000, p. 86-87) (*vide* STERELNY; GRIFFITHS, 1999, p. 220-224).

Psicólogos evolutivos se concentram em adaptações psicológicas. Um tema consistente no trabalho teórico dos psicólogos evolutivos é que "adaptações,

os componentes funcionais dos organismos, são identificados [...] pela [...] evidência de seu design: a combinação extraordinária entre a estrutura do organismo e o ambiente" (HAGEN, 2005, p. 148). O modo no qual adaptações psicológicas são identificadas é através de análise funcional evolutiva, que é um tipo de engenharia reversa.[61] "A engenharia reversa é um processo para compreender o design de um mecanismo com base em uma análise das tarefas que ele realiza. A análise funcional evolutiva é uma forma engenharia reversa na medida em que tenta reconstruir o design da mente a partir de uma análise dos problemas que são tais que a mente tem de ter evoluído para resolvê-los" (BULLER, 2005, p. 92). Muitos filósofos objetam à atribuição excessiva, por parte dos psicólogos evolutivos, de adaptações baseadas em um aparente design. Neste ponto, alguns estão seguindo o exemplo de Gould e Lewontin (1979), ao questionar que explicar o aparente design na natureza em termos de adaptação equivale a contar meras histórias. Mas eles poderiam com igual facilidade citar Williams (1966), que também alertou contra a atribuição excessiva de adaptação para os traços biológicos. Embora seja verdade que a análise funcional evolutiva pode se prestar ao papel de contar meras histórias, este não é o problema mais interessante enfrentado pela psicologia evolutiva. Vários outros problemas interessantes foram identificados. Por exemplo, Elisabeth Lloyd (1999) deriva uma crítica à psicologia evolutiva a partir da crítica de Gould e Lewontin à sociobiologia. Ela enfatiza o ponto de que o adaptacionismo dos psicólogos evolutivos os leva a ignorar processos evolutivos alternativos. Buller adota ainda outra abordagem com respeito ao adaptacionismo dos psicólogos evolutivos. O que está por trás das críticas de Buller ao adaptacionismo dos psicólogos evolutivos é uma perspectiva diferente sobre o que é importante no pensamento evolutivo (BULLER, 2005). Buller pensa que psicólogos evolutivos exageram a importância do design, e fazem a suposição controversa de que, com respeito aos traços nos quais estão interessados, a evolução está terminada e não em andamento.

A definição de Sober de adaptação não tem aplicação limitada a órgãos e outros traços que exibem um aparente design. Pelo contrário, o tamanho da ninhada (em pássaros), os cardumes (em peixes), a disposição das folhas, as estratégias

---

[61] Aqui, seguimos a explicação dada por Buller (2005) desta abordagem. O termo "engenharia reversa" foi primeiramente usado em um contexto evolutivo por Daniel Dennett. (Ele explica e elabora o conceito em seu trabalho de 1995). Assim como muitos outros em psicologia evolutiva, Steven Pinker (1997) também defende esta abordagem.

de forrageamento e todo tipo de traços podem ser adaptações (*vide* SEGER; STUBBLEFIELD, 1996). Buller defende o ponto mais geral de que vários tipos de plasticidade fenotípica podem ser adaptações, porque em vários organismos elas surgem como um resultado da seleção natural.[62] Aqui, a diferença entre Buller (e outros filósofos e biólogos) e psicólogos evolutivos reside no escopo explicativo que eles atribuem à seleção natural. Para psicólogos evolutivos, a característica distintiva da seleção natural é um órgão que funciona bem. Para seus críticos, os resultados da seleção natural podem ser vistos em uma enorme variedade de traços, que vão desde características dos órgãos que aparentemente exibem um design específico até os perfis mais gerais de respostas no comportamento. De acordo com Buller, esta última abordagem abre o leque de possíveis hipóteses evolutivas que podem explicar o comportamento humano. Ao invés de nos restringirmos a explicar nosso comportamento em termos da junção de outputs de muitos mecanismos modulares específicos, podemos explicar nosso comportamento por meio do apelo à ação da seleção em muitos níveis diferentes de traços. Esta diferença de ênfase no que é importante em teoria evolutiva também está no centro de debates entre psicólogos evolutivos e ecologistas comportamentais. Os últimos argumentam que comportamentos, e não apenas os mecanismos subjacentes a eles, podem ser adaptações (*vide* DOWNES, 2001). Além disto, esta diferença de ênfase é o que leva à ampla variedade de hipóteses evolutivas alternativas apresentadas por Sterelny (2003) para explicar o comportamento humano. Dado que filósofos como Buller e Sterelny são adaptacionistas, eles não são críticos do adaptacionismo dos psicólogos evolutivos. Pelo contrário, eles criticam o estreito escopo explicativo do tipo de adaptacionismo adotado pelos psicólogos evolutivos (*vide* DOWNES, 2015).

A crítica de Buller de que psicólogos evolutivos assumem que a evolução está terminada com respeito aos traços nos quais eles estão interessados conecta preocupações sobre a compreensão da teoria evolutiva com preocupações sobre o teste das hipóteses evolutivas. Eis uma afirmação clara, de Tooby e Cosmides, da suposição com a qual Buller está preocupado: "psicólogos evolutivos exploram primariamente o design da arquitetura psicológica e neural, universal e evoluída, que todos compartilhamos em virtude de sermos humanos. Psicólogos evolutivos

---

[62] Peter Godfrey-Smith (1996) examina essa ideia com algum detalhe. Uma versão dela é retomada e defendida em Sterelny (2003).

normalmente estão menos interessados em características humanas que variam devido a diferenças genéticas, porque reconhecem que é improvável que estas diferenças sejam adaptações evoluídas centrais à natureza humana. Dos três tipos de características encontrados no design dos organismos - adaptações, subprodutos e ruído - os traços causados por variantes genéticas são predominantemente ruído evolutivo, com pouca importância adaptativa. Já adaptações complexas provavelmente serão universais nas espécies" (TOOBY; COSMIDES, 2005, p. 39). Esta linha de pensamento também capta a perspectiva dos psicólogos evolutivos acerca da natureza humana: a natureza humana é nossa coleção de adaptações universalmente compartilhadas. Para maior discussão desta explicação da natureza humana, assim como de outras explicações opostas que também possuem base biológica, confira Downes e Machery (2013). O problema aqui é que é falso que adaptações não podem estar sujeitas à variação. O problema subjacente é a noção limitada de adaptação. Adaptações são traços que surgem como resultado da seleção natural, e não traços que exibem design e são universais em uma dada espécie (*vide* SEGER; STUBBLEFIELD, 1996). Como resultado, é perfeitamente consistente argumentar, como Buller faz, que muitos traços humanos podem ainda estar sob seleção, a despeito de poderem ser razoavelmente chamados de "adaptações". Por fim, filósofos da biologia têm articulado vários tipos diferentes de adaptacionismo (*vide* GODFREY-SMITH, 2001; LEWENS, 2009; SOBER, 2000). Alguns destes tipos de adaptacionismo podem razoavelmente ser vistos como impondo restrições ao modo como a pesquisa evolutiva é realizada. Mas o "adaptacionismo explicativo" de Godfrey-Smith (2001) tem um caráter diferente. O adaptacionismo explicativo é a perspectiva de que o aparente design é uma das grandes questões que enfrentamos ao explicar nosso mundo natural, e a seleção natural é a grande resposta (e também a única sustentável) a esta grande questão. O adaptacionismo explicativo é frequentemente adotado por aqueles que querem distinguir o pensamento evolutivo do criacionismo ou design inteligente, e é a forma como psicólogos evolutivos costumam elaborar seu trabalho para distingui-lo do de seus colegas na área mais ampla das ciências sociais. Embora o adaptacionismo explicativo sirva para distinguir a psicologia evolutiva destas abordagens visivelmente diferentes da explicação do design na natureza, ele não impõe muitas restrições claras ao modo como explicações evolutivas devem ser buscadas (*vide* DOWNES, 2015). Os desacordos considerados até agora dizem respeito a diferentes perspectivas sobre a natureza e escopo da explicação evolutiva, mas eles têm ramificações para a discussão sobre o teste de hipóteses.

Se os traços que interessam aos psicólogos evolutivos são universalmente distribuídos, então devemos esperar encontrá-los em todos os humanos. Isto explica parcialmente o peso que psicólogos evolutivos dão a testes psicológicos transculturais (*vide* BUSS, 1990). Quando encontramos evidência de um traço em uma enorme seção cruzada de humanos, corroboramos nossa perspectiva de que o traço é uma adaptação - isto dado a suposição de que adaptações são traços semelhantes a órgãos, produtos da seleção natural que não estão sujeitos à variação. Mas dado a perspectiva mais ampla da evolução defendida por filósofos da biologia, este parece ser um método equivocado para testar uma hipótese evolutiva. Estes testes podem certamente levar ao resultado muito interessante de que certos perfis de preferências são amplamente compartilhados entre culturas, mas eles nada dizem sobre a hipótese evolutiva de que as preferências são adaptações (*vide* LLOYD, 1999; BULLER, 2005).

Outro receio dos críticos da abordagem dos psicólogos evolutivos ao teste de hipóteses é que os últimos dão peso insuficiente a hipóteses alternativas sérias que acomodam os dados relevantes. Buller dedica vários capítulos de seu livro sobre psicologia evolutiva a um exame do teste de hipóteses, e muitas de suas críticas giram em torno da introdução de hipóteses alternativas que se saem igualmente bem, ou até melhor, na explicação dos dados. Por exemplo, ele argumenta que a hipótese do acasalamento preferencial por status se sai melhor na explicação de alguns dos dados dos psicólogos evolutivos relativos à seleção de parceiros do que a hipótese da preferência por status elevado, que é preferida pelos psicólogos evolutivos. Este debate depende dos resultados de testes empíricos. O debate anterior está mais intimamente conectado a questões teóricas em filosofia da biologia.

Eu disse em minha introdução que existe um amplo consenso entre filósofos da ciência de que a psicologia evolutiva é um empreendimento profundamente falho, e alguns filósofos da biologia continuam a nos lembrar disto (*vide* DUPRE, 2012). Entretanto, o consenso em questão não é completo, existem alguns defensores da psicologia evolutiva entre os filósofos da ciência. Um modo de defender a psicologia evolutiva é rebater as críticas. Edouard Machery e Clark Barrett (2007) fazem exatamente isto em sua resenha fortemente crítica ao livro de Buller. Outro modo de defender a psicologia evolutiva é praticá-la (pelo menos na medida em que filósofos podem fazê-lo, ou seja, teoricamente). Isto é o que faz Robert Arp (2006) em um artigo recente. Examinarei brevemente ambas as respostas, abaixo.

Machery e Barrett (2007) argumentam que a crítica de Buller não tem um alvo claro, pois nada há na ideia de que existe uma tradição de pesquisa em

psicologia evolutiva que é distinta do empreendimento mais amplo relacionado à compreensão evolutiva do comportamento humano. Eles argumentam que os preceitos teóricos e métodos são compartilhados por muitos daqueles que trabalham com a biologia do comportamento humano. Por exemplo, muitos são adaptacionistas. Como vimos acima, contudo, tanto os psicólogos evolutivos como os ecologistas comportamentais podem se dizer adaptacionistas, embora a abordagem particular de cada um ao adaptacionismo determine a gama de hipóteses que pode gerar e os traços que podem ser considerados adaptações, além de impactar o modo em que as hipóteses são testadas. Tradições de pesquisa podem compartilhar alguns comprometimentos teóricos gerais e, ainda assim, serem tradições distintas de pesquisa. Em segundo lugar, eles argumentam contra a perspectiva de Buller de que ambientes passados não são suficientemente estáveis para produzir o tipo de adaptações psicológicas proposto pelos psicólogos evolutivos. Eles tomam isto como a afirmação de que nenhuma adaptação pode surgir de uma situação de corrida armamentista evolutiva entre, por exemplo, predadores e presas. Mas, novamente, penso que o desacordo aqui é sobre o que conta como uma adaptação. Buller não nega que adaptações - traços que surgem como um produto da seleção natural - surgem de todos os tipos de ambientes instáveis. O que ele nega é que adaptações com propósitos especiais semelhantes a órgãos são o resultado provável de tais cenários evolutivos.

Arp (2006) defende uma hipótese acerca de um tipo de módulo - visualização de cenários - que é uma adaptação psicológica surgida em nossa história hominídea em resposta à necessidade de fabricação de ferramentas, por exemplo a construção de dispositivos de arremesso de lanças para a caça. Arp apresenta sua hipótese no contexto de uma demonstração da superioridade de sua abordagem à psicologia evolutiva, que ele chama de "Psicologia Evolutiva Estrita", em comparação à "Psicologia Evolutiva Ampla", com respeito à acomodação de evidências arqueológicas e fatos sobre a nossa psicologia. Embora a hipótese de Arp seja inovadora e interessante, de modo algum ele a defende de forma conclusiva. Em parte, isto ocorre porque sua estratégia consiste em comparar sua hipótese com a hipótese de a "fluidez cognitiva" não modular, do arqueólogo Steven Mithen, que foi proposta para explicar os mesmos dados (*vide* MITHEN, 1996). O problema aqui é que a perspectiva de Mithen é apenas uma dentre as muitas alternativas de explicação evolutiva do comportamento humano envolvido na fabricação de ferramentas. Mesmo que a tese modular de Arp seja superior à de Mithen, ele não a comparou com a explicação da fabricação e uso de ferramentas proposta por Sterelny (2003,

2012), ou com a explicação proposta por Boyd e Richerson (*vide* BOYD; RICHERSON, 2005). Consequentemente, ele não exclui a possibilidade de que as últimas sejam alternativas plausíveis. Uma vez que nenhuma das últimas explicações depende da postulação de módulos psicológicos, a psicologia evolutiva não foi adequadamente defendida.

5. Psicologia Moral e Psicologia Evolutiva

Muitos dos filósofos que trabalham com psicologia moral entendem que sua área está sujeita a restrições empíricas. Filósofos adotam duas abordagens principais ao uso de resultados empíricos na psicologia moral. A primeira é recorrer a resultados empíricos (e a teorias empiricamente embasadas da psicologia) para criticar explicações filosóficas da psicologia moral (*vide* DORIS, 2002). A outra é gerar (e, na tradição da filosofia experimental, testar) hipóteses sobre nossa psicologia moral (*vide* NICHOLS, 2004). Para aqueles que pensam que nossa psicologia moral é parcialmente (ou totalmente) baseada em capacidades inatas, a psicologia evolutiva é uma boa fonte de resultados empíricos e de teorias empiricamente embasadas. Uma explicação da constituição de nossa psicologia moral segue da explicação da arquitetura da mente fornecida pela perspectiva da modularidade massiva. Nossos juízos morais são um produto de módulos psicológicos específicos a domínios. Estes módulos são adaptações e surgiram em nossos antepassados hominídeos em resposta a contingências de nossos ambientes (predominantemente) sociais. Atualmente, esta posição é amplamente discutida por filósofos que trabalham com psicologia moral. Segue abaixo um exemplo desta discussão.

Cosmides defende a hipótese em psicologia evolutiva de que dispomos de um módulo de detecção de trapaceiros (*vide* COSMIDES, 1989).[63] Este módulo hipoteticamente subjaz a componentes importantes de nosso comportamento em domínios morais, e se acomoda à perspectiva da modularidade massiva acerca de nossa psicologia em geral. Cosmides (juntamente com Tooby) argumenta que a trapaça é uma violação de um tipo particular de regra condicional que acompanha

[63] A hipótese de Cosmides foi primeiramente gerada e testada no contexto da psicologia do raciocínio, e não da psicologia moral. Ela propôs o módulo de detecção de trapaça para explicar o fato de que nosso desempenho em tarefas de raciocínio melhora drasticamente quanto estas são formuladas em termos de intercâmbio social.

um contrato social. O intercâmbio social é um sistema de cooperação para o benefício mútuo, e trapaceiros violam o contrato social que governa o intercâmbio social (COSMIDES; TOOBY, 2005). A pressão seletiva para um módulo dedicado à detecção de trapaceiros é a presença de trapaceiros no mundo social. O módulo de detecção de trapaceiros é uma adaptação que surgiu em resposta a trapaceiros. A hipótese da detecção de trapaceiros tem sido o foco de um enorme volume de discussões críticas. Contra críticos como Ron Mallon (2008) e Fodor (2008), Cosmides e Tooby (2008) defendem a ideia de que a detecção de trapaças é modular, em contraste com a hipótese de que regras mais gerais de inferência estão envolvidas no tipo de raciocínio por trás da detecção de trapaceiros. Parte da crítica da hipótese da detecção de trapaceiros envolve repetições de críticas da modularidade massiva em geral, enquanto parte trata a hipótese como uma contribuição à psicologia moral e recorre a considerações diferentes. Por exemplo, Mallon (2008) questiona a coerência de abandonar, em nossa concepção de nossa psicologia moral, uma concepção de dever que seja geral a domínios. Esta é uma discussão também em andamento. Consulte Sterelny (2012), para uma seleção de explicações alternativas e não modulares de aspectos de nossa psicologia moral).

6. Natureza Humana

A psicologia evolutiva está bem posicionada para fornecer uma explicação da natureza humana. Como notado acima (seção 1), a psicologia evolutiva tem uma dívida teórica com a sociobiologia humana. E. O. Wilson (1978) recorreu à sociobiologia humana para fornecer uma explicação da natureza humana. Para Wilson, a natureza humana é uma coleção de repertórios comportamentais humanos universais, e estes repertórios comportamentais são melhor compreendidos como produtos da seleção natural. Psicólogos evolutivos argumentam que a natureza humana não é uma coleção de repertórios comportamentais humanos universais, mas uma coleção dos mecanismos psicológicos universais que subjazem a estes comportamentos (TOOBY; COSMIDES, 1990). Como vimos na seção 2, acima, estes mecanismos psicológicos universais são produtos da seleção natural. Tooby e Cosmides colocam esta afirmação como segue: “o conceito de natureza humana é baseado em uma coleção de adaptações psicológicas complexas que são típicas da espécie” (TOOBY; COSMIDES, 1990, p. 17). Assim, para psicólogos evolutivos, “a natureza humana consiste em um conjunto de adaptações psicológicas

presumivelmente universais e únicas dos seres humanos" (BULLER, 2005, p. 423). A explicação nomológica da natureza humana, de Machery (2008), é baseada na, e muito similar a, explicação dos psicólogos evolutivos. Machery diz que "a natureza humana é o conjunto de propriedades que humanos tendem a possuir como resultado da evolução de sua espécie" (MACHERY, 2008, p. 323). Embora sua explicação apele a traços que evoluíram e são universais (comuns a todos os humanos), ela não é limitada a mecanismos psicológicos. Ele considera o bipedismo, por exemplo, parte do aglomerado de traços que formam a natureza humana. A perspectiva de Machery capta elementos da perspectiva da sociobiologia e da psicologia evolutiva acerca da natureza humana. Destas duas explicações, Machery compartilha a ideia de que um traço tem de ser um produto da evolução, ao invés de aprendizado cultural ou enculturação.

Algumas críticas à explicação da natureza humana fornecida pela psicologia evolutiva (e também à explicação nomológica) derivam de preocupações similares àquelas que levaram às críticas da psicologia evolutiva em geral. Na seção 4, vimos que discussões acerca da psicologia evolutiva são fundadas em desacordos sobre como adaptação deveria ser caracterizada, e sobre o papel da variação na evolução. Alguns críticos acusam psicólogos evolutivos de assumir que adaptação não comporta variação. A crítica de Buller (2005) à explicação da natureza humana fornecida pelos psicólogos evolutivos também recorre à variação (*vide* HULL, 1986; SOBER, 1980). A ideia aqui é que, como todos os organismos, os seres humanos exibem grande variação, incluindo variação morfológica, fisiológica, comportamental e cultural (*vide* AMUNDSON, 2000). Buller argumenta que ou a explicação da natureza humana fornecida pela psicologia evolutiva ignora toda esta variação, ou ela falha em explicá-la (*vide* LEWENS, 2015; ODENBAUGH, no prelo; RAMSEY, 2013). Nenhuma explicação que limite a natureza humana a apenas aqueles traços que temos em comum, e que também não estão sujeitos a mudanças, pode explicar a variação humana.

A crítica de Buller (2005) à noção que os psicólogos cognitivos têm da natureza humana (ou à explicação nomológica) se baseia na ideia de que variamos em muitas dimensões, e uma explicação da natureza humana baseada em traços fixos e universais não é capaz de explicar quaisquer destas variações. A ideia de que para explicar a natureza humana devemos explicar a variação humana é apresentada e defendida por psicólogos evolutivos (*vide* BARRETT, 2015), antropólogos (*vide* CASHDAN, 2013) e filósofos (*vide* GRIFFITHS, 2011; RAMSEY, 2013). Barrett concorda com Buller (e outros) que a explicação da natureza humana fornecida

pelos psicólogos evolutivos falha em acomodar a variação humana. Ele não vê isto como um problema decisivo contra o inteiro empreendimento de explicar a natureza humana, e sim como um desafio para uma explicação da natureza humana. Barrett diz que "O que quer que seja a natureza humana, ela é um fenômeno psicológico com tudo que isto implica" (BARRETT, 2005, p. 321). Assim, a natureza humana é "uma grande e instável nuvem que é diferente das populações de nuvens de esquilos e palmeiras. Para compreender as mentes e comportamentos humanos, precisamos compreender as propriedades de nossa própria nuvem, por mais confusa que ela seja" (BARRETT, 2015, p. 232). Para Barrett, a natureza humana não é uma coleção de traços psicológicos universais, fixos e compartilhados, mas sim o inteiro aglomerado de traços humanos, que inclui todas as variações em todos os nossos traços. Esta abordagem da natureza humana é muito diferente das abordagens defendidas por Wilson, Tooby e Cosmides ou Machery, mas ela também está sujeita a diversas críticas. O principal enfoque das críticas é que tal perspectiva não pode ser explicativa, e que, na verdade, ela consiste meramente de uma grande lista de todas as propriedades que os humanos já tiveram e que podem ter (*vide* BULLER, 2005; DOWNES, 2016; FUTUYMA, 1998; LEWENS, 2015). Em muitas das áreas sobre as quais os psicólogos evolutivos se concentram, ainda há discussões sobre a tensão entre as perspectivas dos psicólogos evolutivos e a manifesta variação nos traços humanos. Outro exemplo, desta discussão mais ampla, é incluído na seção 7, abaixo.

7. Aplicações da Psicologia Evolutiva e Prospectos para o Debate Futuro

Recorre-se à psicologia evolutiva em uma vasta gama de áreas de estudo. Exemplos são a literatura inglesa, os estudos do comportamento do consumidor e o direito. Consulte Buss (2005), para uma discussão dos casos da literatura e do direito, e Saad (2007) para uma apresentação detalhada da psicologia evolutiva e os estudos do comportamento do consumidor. A psicologia evolutiva é normalmente introduzida aos profissionais nestes contextos como algo que fornece recursos que levarão a avanços na área relevante. Filósofos têm reagido de forma crítica a algumas destas aplicações da psicologia evolutiva. Um problema é que psicologia evolutiva é frequentemente confundida com evolução ou teoria evolutiva em geral (*vide* LEITER; WEISBERG, 2009; DOWNES, 2013). A discussão analisada na seção 4, acima, revela uma boa dose de desacordo entre teóricos evolutivos e psicólogos evolutivos acerca da explicação adequada da evolução. Psicólogos evolutivos

propõem aprimorar áreas como o direito e os estudos do comportamento do consumidor através da introdução de ideias evolutivas, mas o que é de fato oferecido é uma seleção de recursos teóricos apoiados apenas pelos defensores de uma abordagem específica da psicologia evolutiva. Por exemplo, Gad Saad (2007) argumenta que os estudos do comportamento do consumidor lucrarão muito com o acréscimo do pensamento adaptativo, que consistiria em procurar por um aparente design e introduzir hipóteses de módulos evoluídos para explicar o comportamento do consumidor. Muitos não veem isto como um esforço para aplicar a teoria evolutiva, no sentido mais amplo, aos estudos do comportamento do consumidor (*vide* DOWNES, 2013).

Promover ideias teóricas contestadas é certamente problemático, mas é ainda mais preocupante quando, no esforço de aplicar a psicologia evolutiva, um trabalho completamente desacreditado é promovido. Owen Jones (2000, 2005), que acredita que o direito será beneficiado pela aplicação da psicologia evolutiva, toma como um trabalho evolutivo exemplar a perspectiva amplamente desacreditada de Randy Thornhill e Craig Palmer (2000) de que o estupro é uma adaptação (*vide* DE WAAL, 2000; COYNE; BERRY, 2000; COYNE, 2003; LLOYD, 2003; VICKERS; KITCHER, 2003; KIMMEL, 2003). Jones (2000) ainda afirma que os críticos do trabalho de Thornhill e Palmer não têm credibilidade como cientistas e teóricos evolutivos. Esta afirmação indica uma séria desconexão de Jones com a literatura científica (e filosófica) mais ampla sobre teoria evolutiva (*vide* LEITER; WEISBERG, 2009).

Para além de monitorar os esforços de expansão da psicologia evolutiva, existem diversas outras áreas nas quais um trabalho filosófico adicional em psicologia evolutiva será frutífero. Os exemplos supracitados de trabalhos em psicologia moral mal arranham a superfície desta área, que está em rápido desenvolvimento. Existe um número enorme de hipóteses empíricas que são relevantes para nossa concepção de nossa psicologia moral e que requerem escrutínio filosófico. (Hauser, de 2006, inclui um levantamento de uma ampla variedade de tais hipóteses). Além disto, trabalhos sobre psicologia moral e as emoções podem ser reunidos por meio de trabalhos em psicologia evolutiva e áreas correlacionadas. Griffiths (1997) dirigiu a atenção filosófica para a evolução e as emoções, e Nichols (*vide* NICHOLS, 2004) estabeleceu um contato mais íntimo entre este tipo de trabalho e a psicologia moral. Na filosofia da mente, muito ainda pode ser feito sobre o tema dos módulos. Um caminho que vem sendo percorrido e que pode ser frutífero no futuro diz respeito à integração dos conceitos biológico e psicológico de módulos (*vide* BARRETT; KURZBAN, 2006; CARRUTHERS, 2006). Outra área promissora são os trabalhos

conectando a biologia à psicologia por meio da genética (*vide* MARCUS, 2004). Quanto à filosofia da ciência, não tenho dúvida de que muitas outras críticas à psicologia evolutiva serão apresentadas. Mas uma área de pesquisa filosófica relativamente subdesenvolvida diz respeito às relações entre toda variedade de abordagens teoricamente diferentes à biologia do comportamento humano (*vide* DOWNES, 2005; GRIFFITHS, 2008; BROWN *et al.*, 2011). Psicólogos evolutivos apresentam seu trabalho como estando em conjunto com o trabalho de ecologistas comportamentais, psicobiólogos do desenvolvimento e outros (*vide* BUSS, 2005, 2007), mas eles não confrontam adequadamente as dificuldades teóricas que surgem para um empreendimento integrado na biologia do comportamento humano. Por fim, embora o debate entre cientistas influenciados pela biologia e outros cientistas sociais permaneça vivo, a maioria dos filósofos não prestou muita atenção na potencial integração da psicologia evolutiva ao estudo interdisciplinar mais amplo da sociedade e da cultura (mas veja Mallon e Stich, de 2000, sobre psicologia evolutiva e construtivismo). Em contraste, filósofas feministas prestaram atenção à questão da integração, e também propuseram críticas feministas à psicologia evolutiva (*vide* FEHR, 2012; MEYNELL, 2012; e o verbete sobre **Filosofia Feminista da Biologia**, da SEP). Gillian Barker (2015) compartilha, com os filósofos da biologia discutidos na seção 4, algumas críticas de base evolutiva à psicologia evolutiva. Mas ela também avalia a psicologia evolutiva em relação a outras ciências sociais. Ela contribui com uma consideração crítica inovadora da psicologia evolutiva, argumentando que, tal como atualmente praticada, a psicologia evolutiva não é um guia frutífero para a política social relativa ao florescimento humano.

A publicação da enorme coleção de artigos de Shackleford e Weekes-Shackleford (2017), sobre questões advindas das ciências psicológicas evolutivas, fornece um ótimo recurso para filósofos que procuram por material para alimentar discussões críticas. Muitos psicólogos evolutivos estão cientes das dificuldades que a variação coloca para algumas abordagens estabelecidas em sua área. Como notado acima (seção 6), este problema é enfrentado por aqueles interessados em desenvolver explicações da natureza humana. Mas ele também surge quando nos deparamos com a variação presente em muitos dos comportamentos humanos que psicólogos evolutivos procuram explicar. A agressão humana, por exemplo, varia ao longo de muitas dimensões, e lidar com cada um destes tipos de variação a fim de explicá-los é um desafio para muitos psicólogos evolutivos (*vide* DOWNES; TABERY, 2017). Dado que a psicologia evolutiva é apenas uma dentre as muitas abordagens baseadas na evolução que explicam o comportamento humano, as

discussões críticas mais promissoras sobre a psicologia evolutiva devem continuar vindo de trabalhos que comparam hipóteses extraídas da psicologia evolutiva com hipóteses extraídas de outras abordagens evolutivas e com hipóteses extraídas de outras abordagens das ciências sociais tomadas no sentido mais amplo. Stephan Linquist (2016) adota esta abordagem no que diz respeito ao trabalho dos psicólogos evolutivos sobre culturas de honra. Linquist introduz hipóteses acerca da evolução cultural que parecem ter maior poder explicativo do que aquelas da psicologia evolutiva. Não há dúvida de que o problema mais amplo da tensão entre a psicologia evolutiva e a evolução cultural continuará a atrair a atenção crítica de filósofos (confira Lewens, de 2015, para uma boa e clara introdução e discussão de abordagens alternativas à evolução cultural).

Por fim, não há dúvida de que filósofos da ciência continuarão a checar as credenciais das ideias evolutivas importadas por outras áreas da filosofia. Filósofos da biologia, em particular, ainda levantam a suspeita de que filósofos tomam suas ideias evolutivas da psicologia evolutiva, ao invés de tomarem-nas da biologia evolutiva. Philip Kitcher (2017) expressa este receio com respeito ao apelo que Sharon Street (2006) faz à evolução. Kitcher receia que Street não se baseia "no que é conhecido sobre a evolução humana" (2017, p. 187) ao fornecer uma explicação de como os traços relevantes podem ter emergido. Como notado acima, a noção nomológica de natureza humana, de Machery (2008, 2017), é criticada sob o fundamento de que sua ideia de um traço evoluído advém da psicologia evolutiva, em oposição à biologia evolutiva. Barker (2015) também encoraja os filósofos e cientistas sociais a aproveitar a enorme gama de recursos teóricos que biólogos evolutivos têm a oferecer, ao invés de se concentrarem apenas naqueles fornecidos pelos psicólogos evolutivos.

Referência bibliográfica

ARP, R. The environments of our hominid ancestors, tool-usage and scenario visualization. **Biology and Philosophy**, 21: 95-117, 2006.

BARKER, G. **Beyond Biofatalism***:* Human Nature for an Evolving World. New York: Columbia University Press, 2015.

BARRETT, H. C. **The Shape of Thought**: How mental adaptations evolve. Oxford: Oxford University Press, 2015.

BARRETT, H. C. Evolutionary Psychology, in ed. FRANKISH, W.; RAMSEY, W. M. **Cambridge Handbook of Cognitive Science.** Cambridge: Cambridge University Press, p. 257-274, 2012.

BARRETT, H. C.; KURZBAN, R. Modularity in Cognition: Framing the debate. **Psychological Review**, 113: 628-647, 2006.

BATESON, P. P. G.; MARTIN, P. **Design for a Life**: How behavior and personality develop. London: Jonathan Cape, 1999.

BJORKLUND, D. F.; HERNANDEZ BLASI, C. Evolutionary Developmental Psychology, in ed. BUSS, D. **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, p. 828-850, 2005.

BOOTH, A. **An Evaluation of the Tracking Argument**. MA Thesis, University of Utah, 2004.

BROWN, G.R.; DICKINS, T.E.; SEAR, R.; LALAND, K.N. Evolutionary accounts of human behavioural diversity. **Philosophical Transactions of the Royal Society of London B**, 366: 313-324, 2011.

BULLER, D., 2005, **Adapting Minds**: Evolutionary Psychology and the Persistent Quest for Human Nature. Cambridge, MA: MIT Press.

BUSS, D. International preferences in selecting mates: A study of 37 cultures. **Journal of Cross-Cultural Psychology**, 21: 5-47, 1990.

BUSS, D. **Evolutionary Psychology***:* The New Science of the Mind. Boston: Allyn and Bacon, 2007.

BUSS, D. (ed.). **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, 2005.

BUSS, D. M., HASLETON, M. G. *et al*. Adaptations, Exaptations and Spandrels. **American Psychologist**, 53: 533-548, 1998.

CASHDAN, E. What is a human universal? Human behavioral ecology and human nature, in ed. DOWNES, S. M.; MACHERY, E. **Arguing About Human Nature**. New York: Routledge, p. 71-80, 2013.

CARRUTHERS, P. **The Architecture of the Mind**. Oxford: Clarendon Press, 2006.

COSMIDES, L. The Logic of Social Exchange: Has natural selection shaped how humans' reason? Studies with the Wason Selection Task. **Cognition**, 31: 187-276, 1989.

COSMIDES, L.; TOOBY, J. Neurocognitive Adaptations Designed for Social Exchange, in ed. BUSS, D. **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, p. 584-627, 2005.

COSMIDES, L.; TOOBY, J. Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules

and Detects Cheaters, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology: The Evolution of Morality**: Adaptations and Innateness, (Moral Psychology, volume 1). Cambridge, MA: MIT Press, p. 53-119, 2008.

COWIE, F. **What's Within? Nativism Reconsidered**. New York: Oxford University Press, 1999.

COYNE, J. A. Of Vice and Men: A Case Study in Evolutionary Psychology, in ed. BROWN TRAVIS, C. **Evolution, Gender, and Rape**. Cambridge MA: MIT Press, p. 171-190, 2003.

COYNE, J. A.; BERRY, B. Rape as an Adaptation. **Nature**, 404: 121-122, 2000.

DENNETT, D. C. **Darwin's Dangerous Idea**. New York: Simon and Schuster, 1995.

DE WAAL, F. Survival of the Rapist. **NewYork Times**, Book Review, p. 24-25, April 2, 2000.

DORIS, J. **Lack of character**: Personality and Moral Behavior. Cambridge: Cambridge University Press, 2002.

DOWNES, S. M. Some recent developments in evolutionary approaches to the study of human behavior and cognition. **Biology and Philosophy**, 16: 575-595, 2001.

DOWNES, S. M. Integrating the Multiple Biological Causes of Human Behavior. **Biology and Philosophy**, 20: 177-190, 2005.

DOWNES, S. M. The Basic Components of the Human Mind Were Not Solidified During the Pleistocene Epoch, in ed. AYALA, F.; ARP, R. **Contemporary Debates in Philosophy of Biology**. Oxford: Wiley Blackwell, p. 243-252, 2009.

DOWNES, S. M. Evolutionary psychology is not the only productive evolutionary approach to understanding consumer behavior. **Journal of Consumer Psychology**, 23: 400-403, 2013.

DOWNES, S. M. Evolutionary Psychology, Adaptation and Design, in ed. HEAMS, T; HUNEMAN, P.; LECOINTRE, G.; SILBERSTEIN, M. **Handbook of Evolutionary thinking in the Sciences**. Dordrecht: Springer, p. 659-673, 2015.

DOWNES, S. M. Confronting Variation in the Social and Behavioral Sciences. **Philosophy of Science**, 83: 909-920, 2016.

DOWNES, S.M.; MACHERY, E.. **Arguing About Human Nature**. New York: Routledge, 2013.

DOWNES, S.M.; TABERY, J. Variability of Aggression, in ed. SHACKLEFORD, T. K.; WEEKES-SHACKLEFORD, V. **Encyclopedia of Evolutionary Psychological Science**. New York: Springer, 2017.

DUPRE, J. Normal People. **Social Research**, 65: 221-248, 1998.

DUPRE, J. Review of Steven Pinker, How the Mind Works. **Philosophy of Science**, 66:489-493, 1999.

DUPRE, J. **Human Nature and the Limits of Science**. Oxford: Clarendon Press, 2001.

DUPRE, J. Against Maladaptationism: Or, what's wrong with evolutionary psychology?, in ed. DUPRE, J. **Processes of Life**: Essays in Philosophy of Biology. Oxford: Oxford University Press, pp. 245-260, 2012.

ERESHEFSKY, M. Psychological Categories as Homologies: Lessons from Ethology. **Biology and Philosophy**, 22: 659-674, 2007.

FAUCHER, L. Biophilosophy of Race, in ed. LIVINGSTONE SMITH, D. **How Biology Shapes Philosophy**: New Foundations for Naturalism. Cambridge: Cambridge University Press, p. 247-275, 2017.

FEHR, C. Feminist Engagement with Evolutionary Psychology. **Hypatia**, 27: 50-72, 2012.

FODOR, J. A. **The Modularity of Mind**, Cambridge, MA: MIT Press, 1983.

FODOR, J. A. **The mind doesn't work that way**: the scope and limits of computational psychology. Cambridge, MA: MIT Press, 2000.

FODOR, J. A. Comment on Cosmides andTooby, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Evolution of Morality: Adaptations and Innateness, (Moral Psychology, volume 1). Cambridge, MA: MIT Press, p. 137-141, 2008.

FUTUYMA, D.J. **Evolutionary Biology**. Sunderland, MA: Sinauer, 1998.

GODFREY-SMITH, P. **Complexity and the Function of Mind in Nature**. Cambridge: Cambridge University Press, 1996.

GODFREY-SMITH, P. Three Kinds of Adaptationism, in ed. ORZACK, S. H.; SOBER, E. **Adaptationism and Optimality**. Cambridge: Cambridge University Press, p. 335-357, 2001.

GOULD, S. J.; LEWONTIN, R. The spandrels of San Marco and the Panglossian Paradigm: A critique of the adaptationist programme. **Proceedings of the Royal Society London *B***, 205: 581-598, 1979.

GRANTHAM, T.; NICHOLS, S. Evolutionary Psychology: Ultimate Explanations and Panglossian Predictions, in ed. HARDCASTLE, V. **Where Biology Meets Psychology**. Cambridge, MA: Mit Press, p. 47-56, 1999.

GRAY, R. D.; HEANEY, M. *et al*., Evolutionary Psychology and the Challenge of Adaptive Explanation, in ed. STERELNY, K.; FITNESS, J. **From Mating to Mentality**: Evaluating Evolutionary Psychology, New York: Psychology Press, p. 247-268, 2003.

GRIFFITHS, P. E. Our Plastic Nature, in ed. GISSIS, S. B.; JABLONKA, E. **Transformations of Lamarckism**: From subtle fluids to Molecular Biology. Cambridge, MA: MIT Press, 319-330, 2011.

GRIFFITHS, P. E. Ethology, Sociobiology and Evolutionary Psychology, in ed. SARKAR, S.; PLUTYNSKI, A. **A Companion to Philosophy of Biology**, New York: Blackwell, 393-414, 2008.

GRIFFITHS, P. E. Evolutionary Psychology: History and Current Status, in SARKAR, S.; PFEIFER, J. **Philosophy of Science***:* An Encyclopedia. New York: Routledge, Volume 1, p. 263-268, 2006.

GRIFFITHS, P. E. The Historical Turn in the Study of Adaptation. **British Journal for the Philosophy of Science**, 47: 511-532, 1996.

GRIFFITHS, P. E. **What Emotions Really Are**. Chicago: University of Chicago Press, 1997.

HAGEN, E. H. Controversial issues in evolutionary psychology, in ed. BUSS, D. **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, p. 145-174, 2005.

HAUSER, M. **Moral Minds**: How Nature Designed our universal sense of right and wrong. New York: Harper Collins, 2006.

HAWKES, K. Why do men hunt? Benefits for risky choices, in ed. CASHDAN, E. **Risk and Uncertainly inTribal and Peasant Communities**. Boulder, CO: Westview Press, p. 145-166, 1990.

HEYES, C. M. False belief in infancy: a fresh look. **Developmental Science**: 647-659, 2014a.

HEYES, C. M. Sub-mentalizing: I'm not really reading your mind**. Perspectives on Psychological Science**, 131-143, 2014b.

HIRSHFIELD, L. A.; GELMAN, S. A. **Mapping the mind:** Domain specificity in cognition and culture. New York: Cambridge University Press, 1994.

HRDY, S. **Mother Nature**: Maternal Instincts and How they Shape the Human Species. New York: Ballantine Books, 1999.

HULL, D.L. On Human Nature. **PSA** (Proceedings of the Biennial Meeting of the Philosophy of Science Association), 2: 3-13, 1986.

IRONS, W. Adaptively Relevant Environments Versus the Environment of Evolutionary Adaptedness. **Evolutionary Anthropology**, 6: 194-294, 1998.

JEFFARES, B.; STERELNY, K. Evolutionary Psychology, in ed MARGOLIS, E.; SAMUELS, R.; STICH, S. P. **The Oxford Handbook of Philosophy of Cognitive Science**. Oxford: Oxford University Press, p. 480-502, 2012.

JONES, O. D. Evolutionary Analysis in law: An introduction and application to child abuse. **North Carolina Law Review**, 75: 1117-1242, 1997.

JONES, O. D. An evolutionary analysis of rape: Reflections on transitions. **Hastings Women's Law Journal**, 11:151-178, 2000.

JONES, O. D. Evolutionary Psychology and the Law, in ed. BUSS, D. M. **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, p. 953-974, 2005.

KIMMEL, M. An Unnatural History of Rape, in ed. BROWN TRAVIS, C. **Evolution, Gender, and Rape**. Cambridge MA, MIT Press: 221-234, 2003.

KITCHER, P. Evolution and Ethical Life, in ed. LIVINGSTONE SMITH, D. **How Biology Shapes Philosophy***:* New Foundations for Naturalism. Cambridge: Cambridge University Press, pp. 184-203, 2017.

LALAND, K. N.; BROWN, G. R. **Sense and Nonsense**: Evolutionary Perspectives on Human Behavior. Oxford: Oxford University Press, 2002.

LAUDAN, L. **Progress and Its Problems**. Berkeley: University of California Press. 1977.

LEITER, B.; WEISBERG, M. Why Evolutionary Biology is (so far) Irrelevant to Legal Regulation. **Law and Philosophy**, 29: 31-74, 2009.

LEWENS, T. **Cultural Evolution**. Oxford: Oxford University Press, 2015.

LEWENS, T. Seven Types of Adaptationism. **Biology and Philosophy**, 24:161-182, 2009.

LINQUIST, S. Which evolutionary model best explains the culture of honour? **Biology and Philosophy**, 31:213-235, 2016.

LEWONTIN, R. The evolution of cognition: Questions we will never answer, in ed. SCARBOROUGH, D.; STERNBERG, S. **Methods, Models, and Conceptual Issues**. Cambridge, MA: MIT Press, p. 107-132, 1998.

LITTLE, D. **Varieties of Social Explanation**: An Introduction to the Philosophy of Social Science. Boulder, CO: Westview Press, 1991.

LLOYD, E. A. Evolutionary Psychology: The Burdens of Proof. **Biology and Philosophy**, 14: 211-233, 1999.

LLOYD, E. A. Violence against Science: Rape and Evolution, in ed. BROWN TRAVIS, C. **Evolution, Gender, and Rape**. Cambridge MA, MIT Press, p. 235-262, 2003.

LLOYD, E. A.; FELDMAN, M. W. Evolutionary Psychology: A View from Evolutionary Biology. **Psychological Inquiry**, 13: 150-156, 2002.

MACHERY, E. A Plea for Human Nature. **Philosophical Psychology**, 21: 321-329, 2008.

MACHERY, E. Human Nature, in ed. LIVINGSTONE SMITH, D. **How Biology Shapes Philosophy**: New Foundations for Naturalism. Cambridge University Press, p.204-226, 2017.

MACHERY, E.; BARRETT, H. C. Review of David Buller *Adapting Minds:* Evolutionary psychology and the persistent quest for human nature. **Philosophy of Science**, 73:232-246, 2007.

MALLON, R. Ought we to Abandon a Domain General Treatment of 'Ought'?", in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Evolution of

Morality: Adaptations and Innateness, (Moral Psychology, volume 1). Cambridge, MA: MIT Press, pp. 121-130, 2008.

MALLON, R.; STICH, S. P. The Odd Couple: The compatibility of social construction and evolutionary psychology. **Philosophy of Science**, 67:133-154, 2000.

MARCUS, G. **The Birth of the Mind***:* How a tiny number of genes creates the complexities of human thought. New York: Basic Books, 2004.

MARR, D. **Vision**: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: W.H. Freeman, 1983.

MEYNELL, L. Evolutionary Psychology, Ethology, and Essentialism (Because What They Don't Know Can Hurt Us). **Hypatia**, 27: 3-27, 2012.

MICHEL, G. F.; MOORE, C. L. **Developmental Psychobiology**: An interdisciplinary science. Cambridge: MIT Press, 1995.

MITHEN, S. **The Prehistory of the Mind**: The cognitive origins of art and science. London: Thames and Hudson, 1996.

NICHOLS, S. **Sentimental Rules**: On the natural foundation of moral judgment. Oxford: Oxford University Press, 2004.

ODENBAUGH, J. Human Nature and the Problem of Variation. No prelo.

PINKER, S. **How the Mind Works**. New York: W.W.Norton, 1997.

RAMSEY, G. Human nature in a post-essentialist world. **Philosophy of Science**, 80: 983-993, 2013.

RELLIHAN, M. Adaptationism and Adaptive Thinking in Evolutionary Psychology. **Philosophical Psychology**, 25:245-277, 2012.

RICHARDSON, R. The Prospects for an Evolutionary Psychology: Human Language and Human Reasoning. **Minds and Machines**, 6 (4): 541-557, 1996.

RICHARDSON, R. **Evolutionary Psychology as Maladapted Psychology**. Cambridge, MA: MIT Press, 2007.

RICHERSON, P. J.; BOYD, R. **Not by Genes Alone**: How culture transformed human evolution. Chicago: University of Chicago Press, 2005.

SAAD, G. **The Evolutionary Bases of Consumption**. Mahwah, NJ: Lawrence Erlbaum, 2007.

SAMUELS, R. Evolutionary Psychology and the Massive Modularity Hypothesis. **British Journal for the Philosophy of Science**, 49: 575-602, 1998.

SAMUELS, R. Massively modular minds: evolutionary psychology and cognitive architecture, in ed. CARRUTHERS, P.; CHAMBERLAIN, A. **Evolution and the Human Mind**: Modularity, language and meta-cognition. Cambridge: Cambridge University Press, p. 13-46, 2000.

SAMUELS, R.; STICH, S. *et al*. Reason and Rationality, in ed. NIINILUOTO, I.; SINTONEN, M.; WOLENSKI, J. **Handbook of Epistemology**. Dordrecht: Kluwer, p.131-179, 1999a.

SAMUELS, R.; STICH, S. *et al*. **Rethinking Rationality**: From Bleak Implications to Darwinian Modules, in LEPORE, E.; PYLYSHYN, Z. **What is Cognitive Science?** Blackwell, Oxford, p. 74-120, 1999b.

SEGER, J.; STUBBLEFIELD, J. W. Optimization and Adaptation, in ed. ROSE, M. R.; LAUDER, G. V. **Adaptation**. San Diego: Academic Press, p. 93-123, 1996.

SHAPIRO, L. A.; EPSTEIN, W. Evolutionary Theory Meets Cognitive Psychology: A More Selective Perspective. **Mind and Language**, 13: 171-194, 1998.

SHACKLEFORD, T. K.; WEEKES-SHACKLEFORD, V. (eds.). **Encyclopedia of Evolutionary Psychological Science**, New York: Springer, 2017.

SIMPSON, J. A.; CAMPBELL, L. Methods of evolutionary sciences, in ed. BUSS, D. **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, p.119-144, 2005.

SINGH, D. Adaptive Significance of female physical attractiveness: Role of waist-to-hip ratio**. Journal of Personality and Social Psychology**, 65: 293-307, 1993.

SINGH, D.; LUIS, S. Ethnic and gender consensus for the effect of waist to hip ratio on judgments of women's attractiveness. **Human Nature**, 6: 51-65, 1995.

SINNOTT-ARMSTRONG, W. (ed.). **Moral Psychology**: The Evolution of Morality: Adaptations and Innateness, (Moral Psychology, volume 1). Cambridge, MA: MIT Press, 2008.

SOBER, E. Evolution, Population Thinking and Essentialism. **Philosophy of Science**, 47:350-383, 1980.

SOBER, E. 2000. **Philosophy of Biology**. Boulder, CO: Westview Press.

STANOVICH, K. **The Robot's Rebellion**: Finding meaning in the age of Darwin, Chicago: University of Chicago Press, 2005.

STERELNY, K. The Adapted Mind. **Biology and Philosophy**, 10: 365-380, 1995.

STERELNY, K. **Thought in a Hostile World**: The Evolution of Human Cognition, Oxford: Blackwell, 2003.

STERELNY, K. **The Evolved Apprentice**: How evolution made humans unique, Cambridge, MA: MIT Press, 2012.

STERELNY, K.; GRIFFITHS, P. E. **Sex and Death**: An Introduction to Philosophy of Biology. Chicago: University of Chicago Press, 1999.

STREET, S. A Darwinian Dilemma for Realist Theories of Value. **Philosophical Studies**, 127:109-166, 2006.

THORNHILL, R.; PALMER, C. T. **A Natural History of Rape**: Biological bases of sexual coercion. Cambridge, MA: MIT Press, 2000.

TOOBY, J.; COSMIDES, L. On the universality of human nature and the uniqueness of the individual. **Journal of Personality**, 58: 17-67, 1990.

TOOBY, J.; COSMIDES, L. The Psychological Foundations of Culture, in ed. BARKOW, H.; COSMIDES, L.; TOOBY, J. **The Adapted Mind**. New York: Oxford University Press, p.19-136, 1992.

TOOBY, J.; COSMIDES, L. Conceptual Foundations of Evolutionary Psychology, in ed. BUSS, D. **The Handbook of Evolutionary Psychology**. Hoboken, NJ: Wiley, Hoboken, NJ: Wiley, p. 5-67, 2005.

VICKERS, A. L.; KITCHER, P. Pop Sociobiology: The Evolutionary Psychology of Sex and Violence, in ed. BROWN TRAVIS, C. **Evolution, Gender, and Rape**. Cambridge MA, MIT Press, p. 139-168, 2003.

WILLIAMS, G. C. **Adaptation and Natural Selection**. Princeton: Princeton University Press, 1966.

WILSON, E.O. **On Human Nature**. Cambridge, MA: Harvard University Press, 1978.

YU, D. W.; SHEPARD, G. H. Is beauty in the eye of the beholder? **Nature**, 396: 321-322, 1998.

Agradecimentos

Agradeço a Austin Booth, David Buller, Marc Ereshefsky, Matt Haber, RonMallon, Shaun Nichols e aos pareceristas da Enciclopédia de Stanford pelos úteis comentários sobre os rascunhos deste verbete.

(VI) A Psicologia da Cognição Normativa*

Autoria: Daniel Kelly e Stephen Setman
Tradução: Aluízio Couto
Revisão: L. H. Marques Segundo

Os seres humanos exibem, desde pequenos, uma tendência a identificar, adotar e fazer valer as normas de suas comunidades locais. Normas são as regras sociais que demarcam o que é apropriado, permitido, exigido ou proibido em diferentes situações para os vários membros da comunidade. Essas regras são informais no sentido de que, embora às vezes estejam representadas nas leis formais, tal como a regra que determina em que lado da estrada deve-se dirigir, elas não precisam estar explicitamente codificadas para efetivamente influenciarem o comportamento. Há regras que proíbem o roubo ou a violação de promessas,

---

* KELLY, D.; SETMAN, S. The Psychology of Normative Cognition. *In*: ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy**. Spring Edition. Stanford, CA: The Metaphysics Research Lab, 2021. Disponível em: https://plato.stanford.edu/archives/spr2021/entries/psychology-normative-cognition/. Acesso em: 20 jul. 2021.

The following is the translation of the entry on The Psychology of Normative Cognition by Daniel Kelly and Stephen Setman, 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/spr2021/entries/psychology-normative-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/psychology-normative-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.

mas também há regras que regem qual é a proximidade apropriada para se estar perante uma pessoa enquanto se fala com ela, ou o quão alto se deve falar durante uma conversa. Assim entendidas, as normas regulam um grande âmbito de atividades. Elas exibem variabilidade cultural em suas prescrições e proscrições, embora a sua presença pareça ser culturalmente universal. Algumas normas exibem características frequentemente associadas à moralidade, tais como uma regra aplicada a todos que proíbe causar dano desnecessário. Outras normas se aplicam apenas a algumas pessoas, tais como aquelas que delimitam o vestuário adequado para membros de gêneros diferentes, ou aquelas a respeito das expectativas e responsabilidades atribuídas aos indivíduos que ocupam posições de liderança. As normas prevalecentes em uma comunidade podem ser mais ou menos justas, razoáveis ou imparciais, e podem ser objeto de crítica ou mudança.

Este verbete fornece uma apresentação geral da pesquisa interdisciplinar da capacidade psicológica para a cognição, motivação e comportamento guiados por normas. As noções de norma e normatividade ocorrem em um enorme âmbito de pesquisas, conectando as humanidades e as ciências comportamentais. Pesquisadores primariamente interessados na psicologia característica do comportamento regido por normas adotam o que podemos chamar de abordagens "cognitivo-evolutivas" do assunto. Essas abordagens, comuns nas ciências cognitivas, baseiam-se em uma variedade de recursos e evidências a fim de investigar diferentes capacidades psicológicas. Este verbete descreve como esses recursos e evidências têm sido usados para elaborar explicações dos aspectos mentais cognitivos e motivacionais que ancoram a capacidade de adquirir, obedecer e fazer valer as normas. Ele também descreve como teorias sobre as pressões seletivas e desafios adaptativos, aspectos importantes na evolução humana recente, têm ajudado a informar e a restringir a teorização sobre essa capacidade psicológica, e também como características dessas pressões e desafios podem influenciar a transmissão e evolução cultural das normas.

Para fins de organização, este verbete começa pelo básico e avança adicionando camadas subsequentes de complexidade e detalhe. Os pesquisadores que adotam uma abordagem cognitivo-evolutiva das normas vêm de uma ampla variedade de disciplinas e têm formulado, explorado e debatido posições a respeito de vários aspectos diferentes. Para apresentar um panorama abrangente dessas bibliografias interconectadas, o verbete começa por estabelecer os contornos mais importantes e os princípios centrais: os principais marcos no espaço conceitual comum a diferentes teorias e afirmações. O verbete também fornece uma descrição

mais detalhada dos tipos de recursos teóricos que os pesquisadores têm empregado e identifica dimensões importantes nas quais explicações mais específicas da psicologia das normas têm variado. Ele então examina diferentes fontes de evidências empíricas que já começaram a iluminar outras características filosoficamente interessantes da capacidade para as normas. Por fim, recorrendo a algumas ilustrações que aparecem em debates recentes sobre teoria moral, o verbete termina com uma discussão sobre a relação entre cognição normativa e moralidade.

1. Uma Capacidade Psicológica Dedicada a Normas

Em um grupo de pessoas, normas são as regras que demarcam o que é apropriado, exigido ou proibido para os vários membros do grupo em diferentes situações. Elas tipicamente se manifestam em regularidades comportamentais comuns e são mantidas pelas sanções sociais. Desde pequenos, humanos veem certos comportamentos, contextos e papéis como itens regidos por normas. Tão logo um indivíduo adota uma norma para si, ela opera tanto como uma regra que guia seu comportamento quanto como um padrão contra o qual o comportamento é avaliado. Além disso, indivíduos tipicamente tornam-se motivados a fazer valer as normas que adotam, e assim participam de práticas reguladoras tais como a punição e a atribuição de culpa. Tais práticas, por sua vez, ajudam a estabilizar os arranjos sociais da comunidade e as normas que os estruturam. Normas frequentemente são classificadas em tipos ou subcategorias. Exemplos comuns incluem normas morais, sociais, convencionais, epistêmicas, estéticas e organizacionais. O modo correto ou teoricamente mais útil de distingui-las e de encaixar tipos de normas em taxonomias é objeto de muito debate, mas esse debate será posto de lado aqui (*vide* O'Neill, 2017, para uma revisão e Kelly, no prelo, para uma discussão). Em vez disso, esta seção irá esquematizar um panorama geral do espaço conceitual comum ao trabalho cognitivo-evolutivo a respeito da psicologia das normas. As seções posteriores então detalham seu conteúdo, localizando dentro dele diferentes afirmações e teorias específicas.

Uma ideia crucial para o trabalho sobre a psicologia das normas é a de que mentes humanas contêm algum tipo de **sistema de normas**, um conjunto de mecanismos psicológicos dedicados a lidar com a informação e a produção de comportamentos relevantes para as normas. Tais mecanismos são parte de uma estratégia explicativa comum a toda psicologia (CUMMINS, 2000). Nesse caso, os

teóricos apelam para diferentes propriedades do sistema de normas a fim de ajudar a explicar diferentes aspectos de uma capacidade complexa para o comportamento guiado por normas – uma capacidade para "fazer" normas. Essa capacidade é caracterizada por um padrão amplo, mas ao mesmo tempo distinto, de comportamento: quando expostos a estímulos normativamente relevantes (tipicamente centrados nas ações das outras pessoas ou nas próprias) em conjunto com sinais a respeito do contexto dessas ações e dos papéis dos atores, os indivíduos exibem um tipo de resposta robusta e multifacetada centrada em conformidade e punição. Tomadas em conjunto, as respostas dos indivíduos agregam-se de modo a produzir em nível grupal efeitos estabilizadores nos padrões de organização social coletiva. A complexidade e robustez da capacidade individual sugere a operação de uma maquinaria psicológica dedicada – um sistema de normas – que sensibiliza os humanos em prol de certos estímulos sociais (comportamento, contextos, papéis) e confiavelmente produz as facetas coordenadas (psicologicamente, inferencialmente, comportamentalmente) da resposta característica.

Esse quadro coloca a questão de **como** os humanos são capazes de espontânea e confiavelmente rastrear em seu mundo aspectos importantes relativos às normas, inferir as regras que as regem e efetivá-las no próprio comportamento e no comportamento alheio. Isso também exige uma resposta psicológica que ilumine o que media o estímulo e a resposta. Em uma analogia, um automóvel é capaz de aceleração – ele aumenta sua velocidade confiavelmente em resposta à pressão sobre o acelerador – mas é necessário observar "o que ocorre nos bastidores" mecânicos para ver que tipo de mecanismos estão confiavelmente traduzindo esse tipo de *input* naquele tipo de *output*. Tal como no caso da aceleração do carro, quando o que está em causa é a atividade guiada por normas, abordagens cognitivo-evolutivas postulam e investigam a maquinaria psicológica responsável por traduzir certos tipos de *inputs* sociais em tipos de *outputs* comportamentais associados a normas. Essa estratégia – de postular mecanismos psicológicos que mediam estímulo e resposta – é apoiado agora por um modo familiar de compreensão da mente enquanto um sistema de processamento de informação. Os *inputs* em causa são tratados como informação, que por sua vez é encaminhada e processada por um conjunto de mecanismos psicológicos. Por fim, ela é finalmente traduzida em *outputs* comportamentais. Os estudiosos focados na psicologia particular da cognição normativa têm postulado a existência desse tipo de pacote especializado de mecanismos e também investigado diferentes possibilidades sobre sua natureza.

Introduzindo aqui alguns termos técnicos, as explicações das capacidades cognitivas frequentemente desejam fornecer tanto uma **explicação proximal** quanto uma **explicação distal**. Enquanto as explicações proximais tentam responder "como isso funciona?", explicações distais tentam responder "como viemos a ser dessa maneira?". Essa distinção tem origem na biologia (MAYR, 1961; ARIEW, 2003), mas é também aplicável a traços comportamentais e psicológicos (GRIFFITHS, 2007). Cruciais para as explicações proximais em contextos desse último tipo são os modelos de processos psicológicos subjacentes a capacidades específicas. Após identificar uma habilidade relativamente complexa, tal habilidade é explicada em termos da operação e interação de um conjunto de mecanismos componentes subjacentes relativamente mais simples. Assim, as explicações proximais das normas têm o objetivo de mostrar, por meio da identificação de partes componentes do sistema de normas e da descrição de como elas operam, como indivíduos humanos são psicologicamente capazes de uma rica variedade de atividade associada ao comportamento guiado por normas.

As **explicações distais**, por outro lado, têm o objetivo de explicar as prováveis origens de diferentes traços. Nas ciências cognitivas é agora comum o uso extensivo da teoria da evolução. Considera-se o que é conhecido sobre os ambientes e pressões seletivas com os quais as populações ancestrais se depararam e usa-se essa informação para ajudar a informar hipóteses a respeito da mente (BARKOW; COSMIDES; TOOBY, 1992). Os desafios adaptativos colocados pela ação coletiva e pela cooperação em grande escala são cruciais para muitas das explicações distais propostas pelos pesquisadores interessados em psicologia das normas (GINTIS *et al.* 2005; BOYD; RICHERSON 2005b; HENRICH, N.; HENRICH, J. 2007; TOMASELLO, 2009). Em nome da clareza, é útil lembrar que esses dois estilos de explicação são analiticamente distintos, mas que as explicações proximais e distais para um dado traço idealmente serão complementares e se reforçarão mutuamente. Assim, explicações evolutivas da cognição normativa podem informar e restringir modelos proximais e vice-versa.

O foco psicológico das abordagens cognitivo-evolutivas às normas dá a elas uma agenda de pesquisa bastante clara. É bom notar, porém, que enquanto questões a respeito da natureza das normas são relevantes para uma gama de debates em filosofia (e além dela), o trabalho em psicologia das normas não é primariamente guiado por uma tradição ou debate **filosófico** particular. Em vez disso, os que têm como foco a **psicologia** normativa são em geral motivados por um conjunto de questões mais amplas a respeito da natureza humana: por um lado,

a estrutura e as características distintivas das mentes humanas, os rumos da evolução humana que as produziram, os aspectos em comum e as diferenças entre mentes e comportamentos humanos e, por outro, aqueles encontrados em espécies não humanas (TOMASELLO, 1999; RICHERSON; BOYD, 2005; TOOBY; COSMIDES, 2005; HENRICH, 2015; VINCENT; RING; ANDREWS, 2018). Como resultado, os teóricos recorrem a vários recursos explicativos disponibilizados pela ciência cognitiva contemporânea. Assim, tais tratamentos da cognição normativa não são limitados pelas explicações que a psicologia de senso comum [*folk psychology*] oferece para o comportamento, estando assim livres para postular e recorrer a mecanismos, estados e processos psicológicos que não precisam ter muita semelhança com crenças e desejos, graus de crenças [*credences*] e preferências, deliberação consciente e inferência explícita.

Este verbete está organizado em torno de pesquisas cujo foco é a psicologia característica da cognição normativa. Entretanto, **qualquer** discussão sobre normas e comportamento guiado por normas envolverá, tacitamente ou não, **alguma** perspectiva dos agentes e das características que os tornam responsivos à influência normativa. Algumas começam com formalizações analíticas de tipos de agentes e estados mentais pressupostos pela psicologia de senso comum. Juntamente com vários refinamentos, essas formalizações são usadas para dar conta de diferentes fenômenos relativos às normas. Isso está além do escopo deste verbete, mas veja, especialmente, Bicchieri, Muldoon e Sontuoso (2018) para uma visão geral dessas abordagens (*vide* BICCHIERI, 2006, 2016; BRENNAN *et al.*, 2013; CONTE; ANDRIGHETTO; CAMPENNÌ, 2013; HAWKINS; GOODMAN; GOLDSTONE, 2019; MORRIS *et al*., 2015). É também bom notar que abordagens cognitivo-evolutivas são às vezes apresentadas como algo bem distinto das clássicas abordagens da escolha racional à decisão e comportamento humanos (BOYD; RICHERSON, 2001; HENRICH *et al.* 2001, 2005). Se ambas são alternativas genuinamente distintas é algo que permanece obscuro (ELSTER, 1991; WENDEL, 2001), muito embora aqueles que defendem a distinção tipicamente apontam para um corpo crescente de evidências de que os seres humanos raramente se aproximam dos agentes racionalmente ilimitados e puramente auto-interessados da economia clássica (GIGERENZER; SELTEN, 2001; KAHNEMAN, 2011; MILGRAN, 2019, Outros Recursos da Internet). Para ilustrar, segundos as observações, participantes anônimos em jogos de cooperação de uma etapa cooperaram rotineiramente mesmo quando explicitamente cientes de seu anonimato e do fato de que irão jogar apenas uma vez (MARWELL; AMER, 1981; *vide* Thaler (1992) para uma revisão). Os simpáticos

a abordagens cognitivo-evolutivas das normas têm em mãos um modelo explicativo para essas descobertas, concebendo o comportamento dos participantes como algo motivado por seu sistema de normas e por normas pró-sociais internalizadas por eles.

Com o repertório expandido de entidades psicológicas à disposição, explicações que recorrem à capacidade psicológica dedicada às normas são também, ao que parece, bastante apropriadas para capturar os tipos de dissonância e dissociação que podem ocorrer entre, por um lado, a atenção, categorização implícita e motivação normativa individuais e, por outro, suas crenças explícitas e princípios declarados. Uma pessoa pode, por exemplo, endossar explicitamente o feminismo e sinceramente desejar extinguir as normas e expectativas que nutre sobre as mulheres, mas mesmo assim monitorar o mundo social por meio da lente dessas normas sexistas e experimentar uma motivação persistente de fazer valer essas normas e de obedecê-las (confira o trabalho sobre vieses implícitos de Brownstein e Saul, de 2016). Em suma, nesse tipo de quadro, sistemas psicológicos diferentes que abrangem a mente de um indivíduo (talvez o sistema de normas e o sistema de raciocínio prático) podem operar independentemente e estar ligeiramente desalinhados um com o outro.

Muitos outros aspectos que parecem ser particulares da cognição normativa têm ganhado atenção considerável por parte dos pesquisadores em psicologia. Esses aspectos incluem a propensão a **adquirir** normas, a **obedecê-las** e a **fazê-las valer**. Quando uma pessoa nasce em uma comunidade ou passa a fazer parte de uma, ela precisa ser capaz de identificar e extrair informação sobre a ampla variedade de normas que a molda, a quem e quando diferentes normas se aplicam e quais são as consequências de violá-las. Ela deve ser capaz de enxergar alguns comportamentos **como** normativamente regulados, e então inferir qual é a regra que os rege. O aprendizado de como fazê-lo é às vezes apoiado pelo comportamento pedagógico intencional de seus mentores (STERELNY, 2021), mas não precisa sê-lo (SCHMIDT; RAKOCZY; TOMASELLO, 2011). A aquisição do conhecimento das regras não é o fim da história. Os indivíduos raramente apenas observam essas atividades sociais, mas também passam a participar delas de forma competente. Para tanto, uma pessoa tipicamente aprende a se comportar de acordo com as normas que, segundo ela identifica, aplicam-se a ela; a aquisição dessas normas tem como resultado a pessoa guiar a própria conduta. Por fim, normas prevalentes e padrões de conduta são coletivamente mantidos por uma comunidade quando seus membros as fazem valer, punindo aqueles que falham em segui-los. O ato de fazer valer e a punição são categorias amplas e podem incluir correção, suspensão

de cooperação, comunicação de desaprovação por meio de linguagem corporal ou crítica explícita, ostracização ou mexerico a respeito dos violadores de normas, ou mesmo violência física. Assim, indivíduos tornam-se responsivos às normas e às pressões sociais por meio das quais elas se tornam eficazes, e são motivados a aplicar pressão social aos transgressores.

Perguntas adicionais podem ser feitas a respeito de cada uma dessas propensões. Um conjunto de questões diz respeito aos detalhes da aquisição: que tipo de sinais perceptíveis são salientes em sistemas de normas, induzindo assim uma pessoa a perceber um comportamento, contexto ou papel *como* normativamente regido? E uma vez que uma norma é identificada, qual é a causa de a pessoa internalizá-lo? Talvez o fato meramente estatístico de que a maioria das pessoas se comporta da mesma maneira é suficiente em alguns casos, ao passo que em outros uma resposta em forma de sanção pode ser exigida para ativar o processo de aquisição. Uma segunda família de questões diz respeito à motivação: o comportamento guiado por normas é ele próprio tipicamente guiado por motivações intrínsecas ou instrumentais? As pessoas podem obedecer a uma norma pela própria norma, simplesmente porque sentem que é a coisa certa a se fazer. Elas também podem, no entanto, obedecer a uma norma meramente para evitar punição e vergonha. Alguns comportamentos podem ser guiados por ambos os tipos de motivação. Questões similares e talvez mais intrigantes emergem a respeito das raízes psicológicas da motivação das pessoas para punir os que violam as normas. Uma terceira família de questões pode ser enquadrada em termos daquilo que é inato: em que medida os mecanismos responsáveis pela cognição normativa são inatamente especificados ou culturalmente adquiridos? Além dos mecanismos, será algum conteúdo – algumas das próprias normas – inatamente especificado?

Abordagens cognitivo-evolutivas das normas tratam essas questões como empíricas, e assim valorizam as evidências e almejam ser receptivos a uma ampla gama delas. Antes de examinar mais detalhadamente algumas dessas evidências, será útil familiarizar-nos com os tipos de ferramentas teóricas tipicamente usadas pelos pesquisadores para gerá-las e interpretá-las.

1.1 *Background*: Evolução e Considerações Distais

Uma hipótese de trabalho das abordagens cognitivo-evolutivas é a de que os mecanismos psicológicos subjacentes à cognição normativa são adaptações

evolutivas a importantes pressões seletivas na história da evolução humana (RICHERSON; BOYD, 2005; SRIPADA; STICH, 2007, TOMASELLO, 2009; CHUDEK; HENRICH, 2011; KELLY; DAVIS, 2018; cf. COSMIDES; TOOBY, 1992). Ainda que uma explicação proximal detalhada já estivesse disponível, outras perguntas poderiam ser feitas a respeito da proveniência do sistema de normas: como – isto é, em virtude de quais fatores evolutivos, – as mentes humanas vieram a ser equipadas por esses mecanismos psicológicos? Para qual problema ou problemas adaptativos a cognição normativa era uma solução? Quais pressões seletivas foram primariamente responsáveis pela evolução de um sistema de normas, e qual trajetória filogenética essa evolução tomou? Um breve sumário dos tipos de respostas oferecidas hoje a essas perguntas fornece um contexto útil para a discussão que se segue sobre as explicações proximais.

Um princípio de fundo bastante incontroverso do conhecimento já sedimentado é o de que humanos são animais extraordinariamente sociais e que nossas habilidades hipertrofiadas de aprender e cooperar uns com os outros são cruciais para o que nos distingue de nossos ancestrais primatas mais próximos e de outras espécies de hominídeos. Pensa-se que uma diferença central é a de que as capacidades humanas de imitar a aprender uns com os outros se tornaram fortes o suficiente para sustentar a **cultura cumulativa** (TOMASELLO, 1999; HENDRICH; McELREATH, 2003; LALAND, 2017). A cultura é entendida enquanto informação transmitida entre indivíduos e grupos via comportamento, e não enquanto processos como a transmissão genética (RAMSEY, 2013). Crenças, preferências, normas, habilidades, técnicas, artefatos portadores de informação, etc. são passados de indivíduo para indivíduo, e assim através das populações e entre gerações, principalmente por **aprendizado social** (MATHEW; PERREAULT, 2015). Por exemplo, o desenvolvimento do conjunto de técnicas e habilidades associados ao arremesso de lanças, ou o conhecimento e ferramentas que permitem o uso controlado do fogo, têm sido vinculados ao aumento nas oportunidades de aprendizado social fornecidas por redes sociais expandidas e formas mais complexas de atividade social (THIEME, 1997; GOWLETT, 2006). A cultura é **cumulativa** no sentido de que o corpo de informação no repositório não permanece estático, mas pode crescer e se tornar mais complexo. Cabanas de capim evoluem para casas com estrutura de madeira, e assim para construções de pedras e finalmente para arranha-céus. Líderes tribais evoluem para reis, depois imperadores, depois primeiros-ministros. Conjuntos simples de normas evoluem para institucionais informais mais complicadas, e então para códigos legais formalizados de forma bizantina. À medida em que cada geração acrescenta suas

novas inovações, descobertas e melhorias, a sofisticação funcional é acumulada em traços culturais de modo muito parecido ao modo como ela é acumulada em traços genéticos.

Esse panorama geral evolutivo oferece razões para pensar que à medida que os grupos humanos cresceram em tamanho, também cresceu sua capacidade de sustentar mais cultura e de produzir mais inovações culturais (KLINE; BOYD, 2010; HENRICH, 2015, cap. 12), muito embora a relação causal entre tamanho da população e complexidade cultural permaneça controversa (FOGARTY; CREANZA, 2017; VAESEN *et al.*, 2016). À medida em que as inovações culturais continuaram a se acumular, elas permitiram aos humanos controlar e reformular de modo mais significativo os ambientes nos quais eles viviam. Essas transformações também remodelaram o ambiente habitado por gerações subsequentes, mudando assim os contornos dos nichos físicos, sociais e informacionais nos quais elas evoluíram. Essas mudanças, por sua vez, criaram uma gama de novas pressões seletivas, muitas das quais favoreceram corpos, cérebros e mentes mais bem equipados para a sociabilidade e para a herança cultural. Os pesquisadores continuam a desenvolver e a debater os méritos de diferentes ferramentas conceituais para compreender esse tipo de dinâmica evolutiva (TOMASELLO, 1999; LALAND; ODLING-SMEE; FELDMAN, 2001; LALAND; ODLING-SMEE; MYLES, 2010; STERELNY, 2003, 2012; RICHERSON; BOYD, 2005; TENNIE; CALL; TOMASELLO, 2009; BOYD; RICHERSON; HENRICH, 2011; HENRICH, 2015; BOYD, 2017).

Humanos são capazes de habitar uma ampla variedade de ambientes, e a informação socialmente transmitida – ao contrário de informação inatamente especificada e biologicamente transmitida – é particularmente útil em face da variação ecológica e social (RICHERSON; BOYD, 2013). A informação sobre quais plantas em ambientes diversos são comestíveis e quais são tóxicas tem vantagem adaptativa direta. A informação sobre quais tipos de normas prevalecem nos mais diversos ambientes sociais também é importante, e conhecê-la permite que os indivíduos participem sem atritos em sua comunidade e coordenem-se com outros indivíduos em várias atividades coletivas que vão desde a produção de comida e a criação das crianças até a resposta a ameaças e propostas vindas de forasteiros (CHUDEK; HENRICK, 2011). Embora diferentes tipos de **variantes culturais** possam ser úteis de modos distintos, nem toda a informação socialmente transmissível é igualmente valiosa, e os indivíduos não são aprendizes culturais indiscriminados. Os teóricos postulam que as mentes humanas evoluíram para conter uma quantidade de **vieses** ou **heurísticas** no aprendizado social que ajudam a facilitar o aprendizado mais

seletivo. Esses vieses e heurísticas influenciam quais, em meio às várias variantes culturais às quais o indivíduo é exposto, ele irá de fato adotar para si. Duas heurísticas parecem ser particularmente importantes em amplificar as vantagens de um sistema de herança cultural. Um é o viés de **conformidade**, que induz os indivíduos a adotar aquelas variantes culturais que foram adotadas pela maioria dos membros de sua comunidade (MUTHUKRISHNA; MORGAN; HENRICH, 2016), e o outro é o viés de **prestígio**, que sensibiliza os indivíduos para a hierarquia e o *status*, induzindo-os a moldar seu comportamento com base naqueles que obtiveram sucesso e alta posição social (HENRICH; GIL-WHITE, 2001; CHENG *et al.*, 2012; MANER, 2017). Além desses dois vieses, os pesquisadores postularam outros vieses de aprendizado que podem influenciar a aquisição de normas, incluindo um viés que torna a informação sobre as normas mais fáceis de lembrar do que outras informações, essas não normativas, sobre o comportamento (O'GORMAN; WILSON; MILLER, 2008).

A cultura tem papel de destaque em explicações evolutivas da ultra sociabilidade humana, i.e., a habilidade da nossa espécie de cooperar em escalas notavelmente grandes (TOMASELLO, 2009, 2016; RICHERSON, 2013; veja Hagen e Hammerstein (2006) e Burnham e Johnson (2005) para perspectivas alternativas, e Sterelny, Calcott e Fraser (2013) para um contexto mais amplo a respeito da evolução da cooperação). Uma ideia cada vez mais proeminente é a de que explicar a gama completa de comportamentos envolvidos na sociabilidade humana exigirá que recorramos não apenas à cultura em geral, mas a normas culturalmente transmitidas e instituições em particular (MATHEW; BOYD; VAN VEELEN, 2013). Alguns pesquisadores vêm entendendo que a significância e a complexidade dos desafios adaptativos colocados pela cooperação de grande escala têm implicações para a psicologia humana, argumentando a partir dessas bases que as mentes humanas têm uma capacidade específica para as normas (CHUDEK; ZHAO; HENRICH, 2013) que pode ter evoluído em conjunto com nossas capacidades para a linguagem (LAMM, 2014). Outros argumentam que a seleção cultural de grupo, gerada por várias formas de competição entre grupos culturais como comunidades, tribos, clãs e mesmo nações, contribuíram para a propagação de normas cooperativas mais eficazes (TURCHIN, 2018; RICHERSON *et al.,* 2016, KRASNOW *et al.*, 2015). Segundo essas perspectivas, esses tipos de pressões seletivas remodelaram ainda mais a psicologia social humana, suplementando instintos sociais evolutivos mais antigos com a formação do que tem sido chamado de instintos sociais tribais (RICHERSON; BOYD, 2001; BOYD; RICHERSON, 2008; RICHERSON; HENRICH, 2012). Essa família de "instintos" evolutivos recentes é postulada como algo que

inclui uma capacidade para normas, mas também outras propriedades psicológicas que, pensa-se, refinam o comportamento guiado por normas de várias maneiras, incluindo a sensibilidade a marcadores de pertencimento tribal, a fronteiras entre grupos étnicos (McELREATH; BOYD; RICHERSON, 2003) e a emoções sociais como culpa, orgulho e lealdade.

1.2 Psicologia e Explicações Proximais: Ferramentas Teóricas e Dimensões

A caixa de ferramentas teórica comum às ciências cognitivas e comportamentais têm várias dimensões chave dentro das quais os teóricos fixam seu interesse e exploram posições mais específicas a respeito da psicologia normativa. Capacidades pró-normas são entendidas como complexas no sentido de que são auxiliadas por uma quantidade de processos e subsistemas mais simples e inter-relacionados. Esses sistemas complexos são proveitosamente analisados tendo como referência os sistemas mais simples que os compõem. Quando usados na psicologia, esse método geral de analisar sistemas complexos apelando para o caráter e as interações de suas partes componentes frequentemente assume a forma daquilo que tem sido chamado de funcionalismo homuncular (LYCAN, 1990), associado à doutrina funcionalista sobre a mente. Em uma formulação breve, essa é a perspectiva de que os estados e processos mentais são estados funcionais, identificados com os papéis característicos que eles desempenham no sistema psicológico de que fazem parte (PUTNAM, 1963, 1967; FODOR, 1968; LEVIN, 2018)[64] . Dennett, um pioneiro na defesa do funcionalismo homuncular, encorajou os psicólogos a seguir um método paralelo àquele usado na pesquisa sobre inteligência artificial:

> O pesquisador de IA **começa** com um problema intencionalmente caracterizado (por exemplo, como posso fazer para um computador **entender** perguntas do inglês?), decompõe-no em subproblemas também caracterizados intencionalmente (por exemplo, como faço o computador **reconhecer** perguntas, **distinguir** sujeitos de predicados, **ignorar** análises sintáticas irrelevantes?), e então decompõe esses últimos problemas até que ele chegue a problemas ou descrições de tarefas que são obviamente mecanicistas. (DENNETT, 1978, p. 80)

---

[64] N.T.: A primeira edição do verbete de Levin é de 2004.

Um passo importante em qualquer explicação é delinear o próprio fenômeno alvo. Em ciência cognitiva, esse passo frequentemente assume a forma de caracterizar uma capacidade por meio de análise de tarefa, o que equivale a identificar e distinguir as **tarefas** ou funções que são executadas no exercício da capacidade relevante. Por exemplo, algumas tarefas atualmente tidas como centrais ao funcionamento do sistema de normas incluem aquelas mencionadas acima: aquisição, obediência e a capacidade fazer valer as normas. Outro passo diz respeito à modelagem dos **mecanismos psicológicos** responsáveis por levar a cabo essas tarefas: os algoritmos típicos e os padrões de processamento de informação que executam essas funções. O passo final para completar uma explicação completa de uma capacidade seria uma explicação de como esses mecanismos e algoritmos são **implementados** e, assim, efetivados nas estruturas física, química e biológica e de uma pessoa. Consulte Marr (1982) para a discussão clássica desses níveis diferentes de explicação.

Embora evidências experimentais e outras comportamentais possam ajudar a caracterizar mais diretamente uma capacidade e a identificar as tarefas associadas a ela, muitas questões teoricamente importantes dizem respeito à determinação dos tipos de mecanismo psicológicos que deveriam ser postulados para explicar essas tarefas. Teóricos com perspectivas diferentes podem avançar a hipótese de que há diferentes mecanismos na base de uma capacidade particular ou oferecer explicações distintas a respeito de como um mecanismo executa sua função. Eles podem também concordar ou discordar a respeito de como os mecanismos relevantes estão organizados, desenvolvendo então diferentes explicações para o funcionamento dos algoritmos proprietários desses mecanismos e para os tipos de conexões causais e informacionais que cada mecanismo tem com outros mecanismos e com outros elementos da economia psicológica global de uma pessoa (sistemas perceptuais, memória de curto prazo, sistema de produção de ação, etc.).

Dentro desse espaço conceitual há uma quantidade de dimensões importantes nas quais essas explicações podem variar. A seguinte lista de tais dimensões não é exaustiva, mas fornece uma amostra das que são mais significativas. Os mecanismos psicológicos postulados por diferentes explicações proximais de um sistema de normas podem diferir a respeito de

1. Se e em que medida eles são rápidos, automáticos, intuitivos, não conscientes ou se se encaixam de alguma outra maneira na descrição da cognição de

> "tipo 1", ou são lentos, controlados, demandantes de esforço, conscientes ou se se encaixam de alguma outra maneira na descrição da cognição de "tipo 2".
> **2.** Se e em que medida eles carregam as marcas da modularidade, ou seja, são cognitivamente impenetráveis, informacionalmente encapsulados, específicos ao domínio, etc.
> **3.** Se e em que medida eles exigem ou estão sujeitos a controle voluntário.
> **4.** Se e em que medida os mecanismos e seu conteúdo são aspectos universais da natureza psicológica humana ou, em vez disso, exibem variações entre habitats e culturas,
> **5.** Se e em que medida os mecanismos e seu conteúdo são inatos, adaptações genéticas ou, em vez disso, são aprendidos socialmente e transmitidos culturalmente.
> **6.** Se e em que medida a motivação associada aos mecanismos é intrínseca ou instrumental

A primeira dimensão diz respeito a uma distinção entre processos cognitivos de *tipo 1* e de *tipo 2* feita por teorias de processamento e sistemas duais (*vide* Frankish (2010) para uma revisão; CUSHMAN; YOUNG; GREENE, 2010)). Processos de tipo 1 são tipicamente caracterizados como "rápidos e frugais", processos heurísticos intuitivos que geram respostas "não refinadas-e-imediatas" (GIGERENZER *et al.*, 2000). Esses processos ocorrem automática e inconscientemente e são propensos ao erro, mas compensam em velocidade e eficiência no uso de recursos o que lhes falta em precisão. Processos de tipo 2, em contraste, são tipicamente caracterizados como processos analíticos mais lentos e baseados em regras. Eles requerem mais concentração, esforço cognitivo, ocorrem conscientemente e geram respostas mais precisas.

A segunda dimensão diz respeito à perspectiva de que a mente é, em alguma medida, composta de módulos: mecanismos psicológicos que são informacionalmente encapsulados, razoavelmente autônomos, automáticos e específicos ao domínio (FODOR, 1983; CARRUTHERS, 2006; ROBBINS, 2017).[65]

---

[65] N.T.: A primeira edição do verbete de Robbins é de 2009.

Módulos são informacionalmente encapsulados no sentido de que eles são insensíveis à informação presente na mente mas não contida dentro do próprio mecanismo, fazendo com que seus processos internos não sejam afetados, por exemplo, pelo que a pessoa refletidamente acredita ou prefere. Uma propriedade relacionada é a impenetrabilidade cognitiva. Isso captura o fato de que a informação e os processos dentro do módulo são eles próprios inacessíveis aos sistemas centrais, tais como aqueles envolvidos na introspecção ou deliberação. Embora seja frequentemente possível considerar conscientemente o *output* de um mecanismo modular, os processos endógenos responsáveis por produzi-lo permanecerão opacos à introspecção direta (CARRUTHERS, 2011).

Tratando agora da terceira dimensão, deve estar claro a esta altura que compromissos feitos no interior das duas primeiras dimensões podem dar apoio a diferentes perspectivas a respeito do quanto as expectativas e o comportamento normativamente regidos são suscetíveis ao controle voluntário. Se os processos que servem à capacidade para as normas são em algum grau automáticos, inconscientes e insensíveis a mudanças que a pessoa faz em suas crenças explícitas, juízos ou volições, esses processos seriam capazes de afetar seu comportamento sem a necessidade de qualquer direção de sua vontade e poderiam ajudar a produzir comportamentos e juízos que seriam opostos à vontade. Pesquisadores que se debruçam sobre vieses implícitos podem fornecer recursos úteis para pensar a respeito da relação entre a cognição normativa e o controle voluntário. Trabalhos recentes iluminam quais tipos de estratégias de intervenção são eficazes (LAI *et al.*, 2014; DEVINE *et al.,* 2012) e sugerem que o esforço cognitivo deliberado e o controle voluntário podem, sob certas condições, sobrepor-se à influência da cognição implícita e automática. Sobre a cognição normativa, pesquisas sugerem que o autocontrole pode ser exigido para que se viole uma norma já internalizada, tal como a norma contra a violação de promessas (BAUMGARTNER *et al.,* 2009). Os detalhes, porém, permanecem obscuros (PEACH; YOSHIDA; ZANNA, 2011; YOSHIDA *et al.* 2012, veja também Kelly (no prelo) para discussão sobre as diferenças entre normas internalizadas e normas declaradas).

A quarta e quinta dimensões são onde o debate tradicional inato/adquirido se desenrola com respeito às normas e à psicologia normativa. É parte da concepção padrão sobre os módulos eles serem inatos no sentido de que irão se desenvolver mais ou menos da mesma maneira em humanos normais, independentemente do arranjo cultural. Defensores da assim chamada psicologia evolutiva, que é um modo especialmente conhecido de aplicar o pensamento evolutivo ao comportamento

humano, têm adotado a ideia de módulos e até mesmo argumentado que as mentes humanas são "massivamente modulares", i.e., compostas exaustivamente ou quase inteiramente de mecanismos psicológicos modulares (BARKOW; COSMIDES; TOOBY, 1992; SAMUELS, 1998; CARRUTHERS, 2006). Nessa perspectiva, a psicologia normativa também será modular em vários aspectos. Um modo de desenvolver essa ideia seria por meio do argumento de que todas as culturas humanas são estruturadas por algum conjunto ou outro de normas, sugerindo assim a presença de cognição modular. Os modos em que as normas diferem de um grupo para o próximo podem então ser explicadas recorrendo a um modelo de **cultura evocada** *[evoked culture model*] (TOOBY; COSMIDES, 1992; *vide* também Sperber (1996) para uma explicação diferente da relação entre cognição modular e cultura). De acordo com esse modelo, os comportamentos guiados por normas encontrados através das culturas seriam interpretados como inatamente restritos, enraizados na forma e no conteúdo endógenos das "adaptações cognitivas para a troca social" comuns a todas as mentes humanas. A variação normativa seria explicada por meio do fato de que grupos diferentes vivem em circunstâncias diferentes, e assim a variação nas condições externas com as quais eles se deparam evocam diferentes subconjuntos do conjunto de todas as normas e comportamentos guiados por normas possibilitado pelo sistema de normas. Essa perspectiva tem sido sugerida, mas ainda não está totalmente desenvolvida (*vide* BUCHANAN; POWELL, 2018). Uma família alternativa de perspectivas usa de diferentes maneiras as ideias de inatismo e especificidade ao domínio (FESSLER; MACHERY, 2012). Essas perspectivas, mais desenvolvidas do ponto de vista da psicologia normativa, retratam os humanos como entes dotados de capacidades dedicadas à aquisição e à prática de normas, mas cujos mecanismos subjacentes contêm nenhum ou pouco conteúdo inatamente especificado. Nenhuma norma particular seria, segundo esta perspectiva, inata; ao invés, a capacidade (talvez juntamente com algum conjunto de vieses de aprendizado) guia a aquisição em seu domínio específico e assim equipa os indivíduos com a fácil internalização de seja lá quais forem as normas presentes em seu ambiente social local. (BOYD; RICHERSON, 2005a; SRIPADA; STICH, 2007; CHUDEK; HENRICH, 2011; KELLY; DAVIS, 2018).

Recentemente, uma alternativa geral a esses tipos de perspectivas nativistas e modulares tem sido desenvolvida em mais detalhes. Ela sustenta que os mecanismos psicológicos relevantes para muitas características dos processos de tipo 1 podem ser **dispositivos cognitivos** [*cognitive gadgets*] aprendidos em vez de instintos cognitivos inatos ou módulos. Nessa perspectiva, uma capacidade complexa – ler e escrever ou jogar xadrez, digamos – ainda é ancorada por uma dada quantidade

de mecanismos psicológicos relativamente integrados e processos rotinizados, mas os próprios mecanismos (e não apenas o conteúdo que eles processam) são moldados e empacotados uns com os outros pela evolução cultural. Esses pacotes de habilidades, uma vez disponíveis no repertório cultural do grupo, podem então ser adquiridos pelos indivíduos via processos de aprendizado de domínio geral [*domain-general*] (HEYES, 2018). A ideia de um dispositivo cognitivo fornece uma nova e promissora opção teórica para a psicologia em geral. Seus defensores, no entanto, ainda não lidaram sistematicamente com o problema de determinar se a ideia é a que melhor captura a capacidade para as normas (*vide* STERELNY, 2012, cap. 7) para uma discussão que antecipa essa linha de pensamento).

Outros pesquisadores interessados em cognição moral em um sentido mais geral – o que ultrapassa o trabalho sobre comportamento guiado por normas e inclui trabalhos sobre psicologia do altruísmo, bem-estar, caráter e virtude, emoções morais, ação intencional versus ação não intencional, e assim por diante – têm procurado desenvolver uma analogia entre, por um lado, teorias chomskyanas da aquisição e uso da linguagem e, por outro, a aquisição e aplicação de regras morais (MIKHAIL, 2007, 2011; DWYER; HUEBNER; HAUSER, 2010; HAUSER; YOUNG; CUSHMAN, 2008; ROEDDER; HARMAN, 2010). Essa abordagem é geralmente nativista, postulando uma competência moral universal capaz de guiar o aprendizado especificamente no domínio da moralidade e que contém suficiente estrutura inatamente especificada para dar conta da suposta escassez de estímulos morais com o qual as crianças se deparam quando tentam aprender as normas prevalecentes em seu ambiente local (confira Laurence e Margolis, de 2001, para uma discussão em ciência cognitiva sobre argumentos da escassez de estímulos). Alguns defensores também sugerem que, além de haver informação especificando a estrutura dos mecanismos dedicados à aquisição e processamento de normas morais, algumas normas particulares podem ser elas próprias incluídas como parte da capacidade moral inata; talvez normas contra o incesto ou a causação de dano intencional (por exemplo, MIKHAIL, 2007, 2011). Por outro lado, outros têm criticado essa perspectiva (PRINZ, 2008; STERELNY, 2012), mas só recentemente uma explicação positiva mais detalhada da aquisição de regras começou a ser desenvolvida. Crucial para essa nova e emergente alternativa empirista é a ideia de que os indivíduos são aprendizes racionais de regras, mas se apoiam em estratégias de aprendizado de domínio geral [*general domain*] para adquirir normas a partir de seu ambiente social e não em uma competência moral inatamente especificada de domínio específico [*domain-specific*] (GAUS; NICHOLS, 2017; AYARS; NICHOLS 2017, 2020; NICHOLS, 2021).

A sexta e última dimensão diz respeito à motivação. Especialmente à luz dos papéis que a punição e a recompensa desempenham na estabilização de padrões de comportamento ao nível do grupo, uma ideia inicialmente plausível é a de que a motivação normativa é instrumental (*vide* FEHR; FALK, 2002). Nessas perspectivas, um indivíduo se conforma a uma norma para receber algum benefício, evitar uma repreensão ou porque quer se comportar de um modo que ele pensa que os outros esperam que ele se comporte. Essa motivação seria instrumental no sentido de que as pessoas obedecem a normas meramente como um meio para um fim ulterior que os motiva de modo mais fundamental; em termos contrafactuais, remova a recompensa, a punição e a expectativa social externas que o comportamento obediente às normas por parte do indivíduo desaparecerá. Pesquisadores que exploram esse gênero de explicação têm enfatizado recentemente o papel dos estados psicológicos tais como preferências condicionais, crenças sociais de segunda ordem, i.e., as crenças das pessoas sobre as expectativas dos outros, e as crenças das pessoas sobre as crenças dos outros sobre o que deve ser feito (confira Bicchieri, Muldoon e Suntuoso (2018) para uma discussão sobre essa família de perspectivas).

Outras explicações concebem a motivação normativa como intrínseca (KELLY; DAVIS, 2018; NICHOLS, 2021, especialmente o cap. 10). Nessa perspectiva, uma vez que uma norma é adquirida e internalizada, ela tipicamente se mistura com algum tipo de motivação não instrumental. As pessoas serão motivadas a obedecer e fazer valer uma dada norma como um fim em si, tendo o ímpeto de fazê-lo independentemente das circunstâncias externas ou da probabilidade estimada de que irão receber sanções sociais caso desprezem as normas. A motivação intrínseca não implica, obviamente, conformidade comportamental incondicional. Por exemplo, uma pessoa pode sentir o apelo intrínseco de uma norma que prescreve deixar 20% de gorjeta, mas ainda escolher passar por cima dela e agir motivado pelo autointeresse material, subtraindo assim a quantia do garçom. Essa segunda família de explicações levanta um conjunto mais amplo de questões sobre a natureza psicológica da motivação normativa, e também o problema de determinar se e como ela poderia ser especial. Será a motivação normativa mais bem tratada como uma categoria psicológica própria, primitiva e *sui generis*? Ou será que ela é mais bem concebida como algo gerado por elementos psicológicos mais familiares como desejos, emoções, ímpetos ou outros tipos de estados cognitivos postulados de forma independente, recrutados para trabalhar em conjunção com a psicologia normativa? (*vide* KELLY, 2020).

Uma explicação da psicologia das normas influente e formulada já há algum tempo, oferecida por Sripada e Stich (2007) ilustra como esses tipos de fragmentos

teóricos podem ser reunidos. O modelo preliminar postula dois mecanismos inatos, um *mecanismo de aquisição de normas* e um *mecanismo de execução de normas*. As funções ou tarefas dos mecanismos de aquisição de normas são

a **identificar** sinais comportamentais que indicam a existência de uma norma;
b **inferir** o conteúdo de tal norma, e finalmente;
c **repassar** informação sobre esse conteúdo para um mecanismo de execução de normas.

Por outro lado, as tarefas dos mecanismos de execução de normas são

a **codificar** e **armazenar** essas normas repassadas pelo sistema de aquisição em uma **base de dados composta por normas**, que pode ter alguns processos proprietários para o raciocínio sobre os conteúdos representados ali;
b **detectar** sinais no ambiente imediato que indicam se alguma dessa normas se aplica à situação, e se sim a quem se aplicam;
c gerar motivação para **obedecer** àquelas normas que se aplicam a si;
d gerar motivação para **punir** aqueles que violam as normas que se aplicam a eles.

Sripada e Stich fornecem uma representação pictórica inicial:

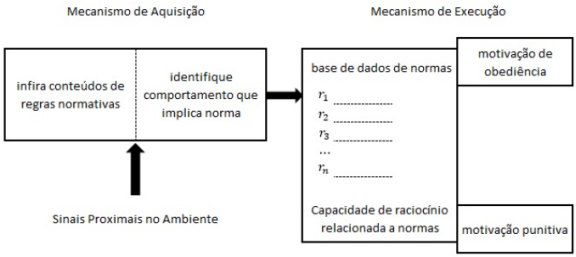


FIGURA: Sripada e Stich (2007, p. 290, figura redesenhada).
Uma descrição estendida da figura está no suplemento.

Postula-se que os mecanismos de aquisição e execução sejam inatos, mas também altamente sensíveis ao arranjo social no qual um indivíduo se desenvolve. Como descrito acima, pensa-se que essa bifurcação entre, por um lado, arquitetura psicologicamente inata e, por outro, conteúdo normativo socialmente aprendido, explica por que a presença das normas é culturalmente universal, ao passo que comportamentos, papéis e arranjos sociais regidos por essas normas apresentam variação. Além disso, o modelo retrata a operação de vários componentes do sistema de normas como "automáticos e involuntários" (SRIPADA; STICH, 2007, p. 290), mas não assume posição sobre processos particulares ou características mais granulares associados à modularidade ou ao processamento dual. Por fim, o modelo é projetado para acomodar evidências que sugerem que quando uma norma é adquirida e representada na base de dados, ela ganha assim um tipo distinto de perfil motivacional. Especificamente, esse perfil concebe a motivação normativa como

1 intrinsecamente motivadora e não instrumentalmente motivadora
2 orientada tanto para si quanto para o outro
3 potencialmente poderosa

Nessa perspectiva, a motivação normativa tem a terceira propriedade no sentido de que em alguns casos ela é capaz de derrotar até mesmo motivações bem convincentes que instam o agente em outras direções; exemplos extremos incluem homens-bomba derrotando seus instintos em prol da autopreservação, ou outros fanáticos que despendem recursos significativos para forçar suas normas sobre os outros. O direcionamento tanto a si quanto aos outros da motivação normativa captura a ideia de que o sistema de normas produz motivação para manter o próprio comportamento sujeito às normas e também motivação para fazê-las valer por meio da punição a quem as viola.

Por fim, o modelo retrata a motivação normativa como intrínseca no sentido comum de que as pessoas obedecem a normas enquanto fins últimos, ou em si mesmas. Sripada e Stich sugerem que a motivação intrínseca ajuda a explicar uma propriedade das normas que eles chamam de "normatividade independente". Isso demarca o fato de que normas podem exercer sólida influência no comportamento das pessoas mesmo quando não estão escritas ou formalmente articuladas em qualquer instituição formal, não sendo assim levadas a cabo por qualquer mecanismo

oficial de punição e recompensa (*vide* DAVIDSON; KELLY, 2020). Eles também discutem a motivação e a normatividade independente em termos de uma "hipótese de internalização" baseada na sociologia e na antropologia, e sugerem que a ideia de internalização pode ser interpretada nos termos de seu modelo. Nessa perspectiva, uma pessoa internalizou uma norma quando a norma foi adquirida por e representada em sua base de dados normativa. A hipótese da internalização pode então ser concebida como a afirmação segundo a qual normas internalizadas são intrinsecamente motivadoras. Isso seria assim por uma simples razão: um aspecto psicológico fundamental da psicologia normativa é que, uma vez que uma norma foi adquirida, entregue à base de dados de normas de uma pessoa e representada nessa base, o sistema de normas automaticamente confere à norma esse caráter distintivamente motivacional. Estar acompanhada por motivação intrínseca direcionada tanto a si quanto aos outros é parte do papel funcional que uma regra vem a ocupar uma vez representada na base de dados de um sistema de normas de uma pessoa – quando é "internalizada" – algo analogamente ao modo como estar acompanhado pela repulsa e pela sensibilidade à contaminação é parte do papel funcional que um sinal vem a ocupar uma vez representado no sistema de aversão [*disgust system*] (KELLY, 2011; *vide* GAVRLILETS; RICHERSON, 2017, para um modelo computacional que explora a evolução da internalização das normas e dos tipos de forças seletivas que podem ter dado à psicologia normativa essa característica intrigante).

2. Pesquisa Empírica

As estratégias explicativas e a caixa de ferramentas teórica das ciências cognitivas têm sido usadas para guiar e explicar uma enorme gama de trabalho empírico. As abordagens cognitivo-evolutivas da psicologia normativa também são interdisciplinares e buscam acomodar a pesquisa empírica feita por antropólogos, sociólogos, economistas comportamentais e psicólogos desenvolvimentistas, comparativos e de outros tipos. Esta seção oferece uma amostra dos tipos de descobertas que vêm sendo reunidas para iluminar aspectos interessantes do comportamento guiado por normas e também para apoiar diferentes posições sobre a cognição normativa.

2.1 Sociologia, Antropologia e Psicologia Cultural

Como ressaltado anteriormente, o registro etnográfico indica que todas as culturas são estruturadas em torno de normas – regras que guiam o comportamento e os padrões a partir dos quais o comportamento é avaliado (BROWN, 1991). As evidências também sugerem que, de uma perspectiva evolucionista, normas são bem antigas – não há muitos indícios de que a capacidade para as normas tenha se disseminado de sociedade para sociedade no passado recente. Os antropólogos também mostraram que normas que regem, por exemplo, o compartilhamento de comida, as práticas conjugais, as redes de parentesco, os rituais comunais, etc., regulam as práticas de atuais caçadores-coletores e de grupos relativamente culturalmente isolados, o que seria improvável se as normas fossem uma motivação recente (*vide* J. HENRICH, 2015). Bastante atenção, entretanto, tem sido dada aos modos pelos quais os conjuntos dominantes de normas variam entre culturas (HOUSE *et al*., 2020; HOFSTEDE, 1980, 2001) e ao modo pelo qual pacotes de normas se desenvolvem e mudam ao longo do tempo dentro de culturas particulares (GAUS, 2016; SCHULZ *et al.*, 2019; INGLEHART, 1997; BEDNAR *et al.*, 2010).

Por exemplo, uma linha de evidências da etnografia comparativa tem como foco o comportamento cooperativo. Ela revela variação entre grupos até mesmo em tipos de atividades, relações e contextos que são regidos por normas. Alguns grupos "cooperam apenas na guerra e na pesca, ao passo que outros grupos, logo rio-abaixo, cooperam apenas na construção de casas e em rituais comunais" (CHUDEK; ZHAO; HENRICH, 2013, p. 426). Que variações comportamentais como essa possam persistir mesmo em face do mesmo contexto ecológico (i.e., "logo rio-abaixo") sugere que elas se devem a diferenças nas normas e em outros elementos socialmente construídos da cultura, não sendo assim respostas mais diretamente evocadas pelo ambiente físico (*vide* HENRICH, N.; HENRICH, J., 2007).

É uma platitude que normas individuais diferentes, identificadas pelo contexto em que se aplicam, por seu escopo, conteúdo, e pelos comportamentos específicos que prescrevem e proscrevem, estão presentes em culturas distintas. Trabalho empírico sistemático tem investigado recentemente a proeminência de diferentes **temas** normativos através das culturas. Exemplos conhecidos incluem diferentes famílias de normas que demarcam culturas da honra versus culturas da vergonha, especialmente aquelas que regem a violência e sua repercussão (NISBET; COHEN, 1996, USKAL *et al.*, 2019), ou os diferentes tipos de normas encontradas em sociedades que premiam valores individuais em detrimento daquelas dominadas

por valores mais coletivistas, especialmente normas que delimitam o escopo da escolha pessoal (McAULIFFE *et al.* 2003; NISBETT, 2004; ROSS, 2012; HAGGER; RENTZELAS; CHATZISARANTIS, 2014; HENRICH, J. (no prelo); veja também Henrich, Heine e Norenzayan (2010) para uma discussão sobre questões metodológicas). Outros pesquisadores têm feito distinções temáticas adicionais ao separar, por exemplo, os tipos de valores e normas de "pureza" que predominam em uma comunidade regida pelo que eles chamam de ética da divindade dos que são prevalentes em comunidades regidas por uma ética da autonomia ou uma ética da comunidade (SHWEDER *et al.*, 1997; ROZIN *et al.*, 1999). Essa linha de pensamento tem sido ainda mais desenvolvida na influente Teoria dos Fundamentos Morais (HAIDT, 2012; GRAHAM *et al.*, 2013). Os teóricos também recorrem a esses tipos de descobertas empíricas para ajudar a avaliar afirmações sobre a psicologia das normas, lançando luz naquelas características da cognição normativa individual que são mais rígidas e universais versus aquelas mais culturalmente maleáveis, e esclarecendo também como esses aspectos psicológicos podem tornar vários padrões de variação grupal mais ou menos prováveis (O'NEILL; MACHERY, 2018).

Uma contribuição recente e intrigante nessa linha é a investigação de Gelfand e colegas a respeito dos padrões de **rigidez** ou **frouxidão** de diferentes normas culturais. Esse trabalho tem como foco diferenças gerais na "força" das normas dentro e através das culturas: quantas normas há, quão tolerantes os membros de uma cultura tendem a ser com desvios do comportamento normativamente prescrito, e quão severamente eles punem violações (GELFAND; NISHII; RAVER 2006; GELFAND *et al.*, 2011; GELFAND; HARRINGTON; JACKSON, 2017). Culturas mais rígidas têm padrões mais numerosos e exigentes, com membros menos tolerantes aos menores desvios e prontos a impor sanções mais severas. Culturas cujos membros são mais lenientes e aceitam maior flexibilidade a respeito de uma norma (sendo assim menos extremos quando fazem-nas valer) recaem na ponta mais frouxa do espectro. Gelfand e colegas exploram as manifestações da estreiteza e frouxidão não apenas ao nível das culturas, mas também ao longo de outros níveis de descrição, do comunal e histórico até o comportamental, cognitivo e neural (GELFAND, 2018). Um ponto crucial para essa perspectiva é o de que a orientação da cultura perante as normas – se os sistemas de normas de seus membros tendem a ser calibradas de modo mais rígido ou frouxo – reflete a severidade dos desafios encarados em seu passado e presente:

> [a] evolução da força da norma é **adaptativa** relativamente a aspectos do ambiente ecológico e, assim, é possibilitada por um conjunto de processos adaptativos psicológicos. (GELFAND; HARRINGTON; JACKSON, 2017, p. 802, ênfase nossa)

Um grupo cuja ecologia é caracterizada por coisas como desastres naturais frequentes, doenças, invasão territorial ou escassez de recursos tem mais chances de ter um sistema de normas mais abrangente e exigente e de assumir uma posição mais rígida a respeito de suas normas, em parte porque a ação social coordenada de modo eficiente é algo exigido para superar ameaças severas. Grupos que se deparam com fatores ecológicos menos estressores são menos forçados à necessidade de ação social rigidamente coordenada, e assim têm condições de dispor de normas mais fracas e uma maior tolerância ao desvio.

2.2 Economia Comportamental

O ponto de partida da economia comportamental é o foco em como pessoas reais tomam decisões econômicas reais e na explicação do tipo de processamento de informação que as leva a não corresponder à racionalidade econômica ideal (KAHNEMANN, 2011). Nas últimas décadas, muitos economistas comportamentais também começaram a investigar variações transculturais no comportamento econômico e a interpretar essas descobertas em termos, por exemplo, das diferentes normas de justiça, equidade e cooperação adotadas pelos participantes (LESOROGOL, 2007). Muita dessa evidência vem de padrões de como as pessoas em culturas diferentes se comportam em jogos econômicos (HENRICH *et al.*, 2001, 2005). Essa variação, por exemplo, foi encontrada em experimentos como o jogo do ultimato, em que dois participantes negociam a respeito de como dividir uma soma não-trivial de dinheiro. O primeiro participante faz uma proposta de como dividir a soma entre os dois, que é então oferecida como um ultimato para o outro. O segundo participante pode aceitar ou rejeitar a oferta. Se ele aceitar, ambos recebem as respectivas quantias especificadas pela proposta; se ele rejeitar, no entanto, ambos os participantes nada recebem. Se eles fossem agentes econômicos ideais, o primeiro participante, agindo de forma autointeressada, ofereceria a menor quantidade possível diferente de zero para o segundo participante, que então aceitaria a proposta porque algo é

melhor que nada. Esse resultado é bastante raro em humanos, no entanto (embora seja comum, o que é intrigante, em chimpanzés, confira Jensen, Call e Tomasello, 2007). Pessoas reais não apenas divergem disso, mas o fazem de maneiras distintas. Vários experimentos descobriram que fatores culturais afetam como as pessoas tendem a jogar o jogo e que a variabilidade nas normas e nas concepções de justiça podem ajudar a explicar diferentes padrões nos tipos de ofertas que os participantes fazem e que estão dispostos a aceitar (ROTH *et al.*, 1991; *vide* OOSTERBEEK; SLOOF; VAN DE KUILEN, 2004) para discussão sobre dificuldades em interpretar esses resultados). Outros experimentos usam um âmbito maior de jogos a fim de reunir evidências de padrões culturais similares de variação cultural no comportamento econômico (*vide* HENRICH *et al*., 2004, para uma coleção desses trabalhos).

Outra família de descobertas, intrigante do ponto de vista da racionalidade econômica clássica, mostra que indivíduos irão rotineiramente punir uns aos outros mesmo quando isso acarreta custos para eles próprios (FEHR; GACHTER, 2002; HENRICH *et al.,* 2006). As evidências sugerem que essa propensão a punir é influenciada por normas e também por outros fatores culturais (BONE; McAULIFFE; RAIHANI, 2016). Por exemplo, em jogos de bens públicos, os participantes recebem somas não triviais de dinheiro e têm, ao longo de várias rodadas, de decidir se e quanto contribuir para uma reserva comum. Quanto o investimento compensa depende de quanto todos coletivamente contribuem na rodada, e assim as decisões de cada participante devem influenciar o comportamento de todos os outros. Em algumas versões, os participantes podem também usar o dinheiro para punir os outros, e o fazem baseados no conhecimento de quanto estes últimos contribuíram. Os resultados indicam que alguns participantes estão dispostos a incorrer em custos para si mesmos para sancionar os que dão contribuições baixas, mas também para punir quem oferece contribuições *altas*, um fenômeno surpreendente chamado *punição antissocial*. Participantes de diferentes culturas exibem padrões diferentes em sua disposição para punir os outros, incluindo seu entusiasmo para a punição antissocial (HERRMAN; THÖNI; GÄCHTER, 2008). Outros aspectos notáveis do comportamento punitivo revelado pelos experimentos em economia comportamental é o de que os humanos estão dispostos a punir mesmo quando eles são meros espectadores do incidente ao qual reagem. Em casos de *punição por terceiros*, um indivíduo faz valer uma norma a despeito do fato de ele não ser nem o ofensor que comete a violação, tornando-se assim alvo da punição, nem o ofendido que foi prejudicado ou a vítima afetada pela transgressão (FEHR; FISHBACHER, 2004; BONE; SILVA; RAIHANI, 2014).

De modo mais amplo, a propensão psicológica geral para a punição iluminada por esse tipo de trabalho empírico emerge, segundo têm dito os pesquisadores, bem cedo em humanos (SCHMIDT; TOMASELLO, 2012; McAULIFFE; JORDAN; RICHERSON, 2015). Alguns têm argumentado que isso é crucial para um conjunto de aspectos da vida social humana, incluindo a estabilização das normas ao nível do grupo (BOYD; RICHERSON, 1992) e a capacidade de sustentar a cooperação em larga escala (PRICE; COSMIDES; TOOBY, 2002; MATHEW; BOYD, 2011; MATHEW; BOYD; van VEELEN, 2013). Outros têm usado esses resultados para fornecer apoio a inferências sobre o caráter da psicologia normativa, incluindo a natureza da motivação normativa. Por exemplo, Chudek e Henrich sumarizam vários estudos neuroeconômicos (FEHR; CAMERER, 2007; TABIBNIA; SATPUTE; LIEBERMAN, 2008; e QUERVAIN *et al.*, 2004) que investigam o comportamento econômico usando métodos e tecnologias da neurociência (ou seja, fMRI) ao apontarem que

> [t]anto cooperar quanto punir em modos localmente normativos ativam os circuitos cerebrais de recompensa ou de antecipação de recompensa da mesma maneira que obter um pagamento em dinheiro. (CHUDEK; HENRICH, 2011, p. 224)

2.3 Psicologia Desenvolvimentista e Comparativa

Uma gama impressionante de evidências sugere que seres humanos são, de nascença, aprendizes de normas. A trajetória do desenvolvimento da cognição guiada por normas em humanos parece exibir similaridades robustas entre diferentes culturas, com as crianças iniciando sua participação no comportamento normativo por volta da mesma idade, que é baixa (*vide* HOUSE *et al.*, 2013; TOMASELLO, 2019). Entre três e cinco anos de idade as crianças exibem o conhecimento de diferentes tipos de regras normativas (TURIEL, 1983; SMETANA, 1993; NUCCI, 2001), e já aos três anos de idade são capazes de executar de forma competente tarefas de raciocínio deôntico (CUMMINS, 1996; BELLER, 2010). Elas fazem valer as normas quando acreditam que o comportamento normativo foi livremente escolhido (JOSEPHS *et al.*, 2016), mas também quando entendem que ele não foi intencional (SAMLAND *et al.*, 2016) ao menos em algumas circunstâncias (CHERNYAK; SOBE,

2016; *vide* BARRETT *et al.*, 2016; CURTIN *et al.*, 2020) para evidência e discussão sobre a variação transcultural da sensibilidade das pessoas perante estados mentais dos violadores de normas). Além disso, crianças são alertas a como as outras pessoas respondem a transgressões, exibindo mais sentimentos positivos em relação àqueles que respondem a uma violação das normas do que em relação aos que permitem que as violações não sejam corrigidas (VAISH *et al.,* 2016).

Talvez o aspecto mais impressionante seja a facilidade e a rapidez com que as crianças adquirem normas. Descobriu-se que pré-escolares aprendem normas rapidamente (RAKOCZY; WARNKEN; TOMASELLO, 2008) mesmo sem instruções explícitas (SCHMIDT; RAKOCZY; TOMASELLO, 2011), embora o aprendizado seja facilitado quando as normas são modeladas por adultos (RAKOCZY *et al.*, 2010). O entusiasmo das crianças por regras, sua "promiscuidade normativa" (SCHMIDT *et al.,* 2016), parece até superar a sensibilidade ao comportamento comum regido por normas em seu ambiente social. Evidências sugerem que às vezes uma única observação de uma ação é suficiente para as crianças inferirem a existência de uma norma, e que, deixadas por conta própria, elas espontaneamente irão criar suas próprias normas e ensiná-las aos outros (GÖCKERITZ; SCHMIDT; TOMASELLO, 2014).

Isso dito, comportamentos tidos como normais em uma comunidade são particularmente salientes à psicologia normativa dos indivíduos. As crianças também são normativamente promíscuas no sentido de que parecem ser propensas a falsos positivos no curso de aquisição de normas, encarando um comportamento como se ele fosse guiado normativamente mesmo quando ele é apenas comum, e inferindo a presença de regras normativas quando elas não existem. Uma série de estudos descobriu que quando participantes (crianças e adultos dos Estados Unidos e da China) detectaram ou foram comunicados que um tipo de comportamento era comum entre um grupo de pessoas, eles passaram a avaliar negativamente membros do grupo que se comportavam de maneira desviante (ROBERTS *et al.*, 2018; ROBERTS; HO; GELMAN, 2019). Os pesquisadores têm investigado essa característica da aquisição das normas a partir de diferentes ângulos e a rotulado com vários nomes, incluindo "tendência-da-descritividade-para-normatividade" (ROBERTS; GELMAN; HO, 2017), "heurística moral comum" (LINDSTRÖM *et al.*, 2017), e "hipótese da inferência injuntiva" (DAVID; HENNES; RAYMOND, 2018, discutindo, e.g., Schultz *et al.* (2007)). Uma vez que as evidências sugerem que que a cognição normativa humana convida a uma fácil inferência de um "é" a respeito de um padrão percebido de comportamento comum para um "deve" relativo a uma norma (TWOREK;

CIMPIAN, 2016), os filósofos podem ficar tentados a chamar tal coisa de “viés da falácia naturalista”.

Outra linha de pesquisa sugere que um aspecto crucial para entender as raízes da normatividade humana é o fato de que crianças são **super imitadores**. Elas não são apenas imitadoras espontâneas, intuitivas e excelentes, mas também tendem a copiar **todos** os elementos da sequência do comportamento de um modelo, mesmo quando elas reconhecem que alguns desses elementos são supérfluos para tarefa em causa (LYONS; YOUNG; KEIL, 2007; KENWARD; KARLSSON; PERSSON, 2011; KEUPP; BEHNE; RAKOCZY, 2013; NIELSEN; KAPITÁNY; ELKINS, 2014; HEYES, 2018, cap. 6). As crianças se atentam à maneira específica pela qual uma ação é levada a cabo (não meramente ao objetivo ao qual ela é direcionada) e assim se conformam a todo o *script* mesmo quando percebem que o objetivo poderia ser atingido de uma maneira mais direta. Além disso, as crianças monitoram os outros para ver se eles estão agindo da mesma forma, e impõem a super imitação sobre seus pares por meio da crítica àqueles que não executam a sequência inteira de passos (KENWARD, 2012; RAKOCZY; SCHMIDT, 2013). A super imitação pode levar a gasto desnecessário de energia dedicado a esses comportamentos irrelevantes, mas o traço pode, entretanto, ser uma adaptação. De acordo com esse argumento, os custos do que parecem “erros” individuais são, em última instância, compensados pelos benefícios comunais gerados por uma população cujas capacidades individuais de transmitir normas e outras variantes culturais são mais persistentes dessa maneira, fazendo com o erro seja em favor do excesso de imitação, e não da falta (HENRICH, cap. 7). Seja lá a razão pela qual tenha sido inicialmente selecionada, os pesquisadores têm sugerido que a maquinaria psicológica selecionada responsável pela super imitação oferece importantes contribuições à cognição normativa. As evidências indicam que essa maquinaria gera forte (e talvez intrínseca) motivação social direcionada à conformidade comportamental com os outros. Quando trabalha em conjunto com um sistema de normas, essa fonte de motivação pode também facilitar a performance da tarefa crucial de manter o comportamento de um indivíduo sob controle, induzindo-o a se conformar não apenas aos comportamentos que ele observa, mas também àquelas normas que ele internalizou (*vide* HOEHL *et al*., 2019).

A super imitação é também notável por poder ser distintivamente humana. Por exemplo, embora chimpanzés imitem o modo como seus coespecíficos manipulam instrumentalmente seu ambiente para alcançar um objetivo, eles copiam o comportamento de modo apenas seletivo, pulando passos que reconhecem como desnecessários (WHITEN *et al.*, 2009; *vide* CLAY; TENNIE, 2018). As evidências

sugerem que o aprendizado em crianças humanas é comparativamente mais sensível a influência dos pares também de outras maneiras. Uma vez que os chimpanzés e orangotangos entendem como resolver um problema, eles se mostram conservadores, fiando-se a qualquer solução que aprenderam primeiro. Os humanos, em contraste, irão com frequência alternar para uma nova solução demonstrada por seus pares, às vezes até mesmo alternando, sob influência daqueles, para estratégias **menos** eficazes (HAUN; REKERS; TOMASELLO, 2014).

Entretanto, outros pesquisadores têm recentemente contestado a afirmação de que a super imitação é estritamente ausente em não humanos (ANDREWS, 2017). Esse é um *front* de um debate muito mais amplo que diz respeito a quais aspectos da psicologia humana são únicos à nossa espécie e quais são compartilhados com outras espécies. Pesquisa recente e relevante para a discussão sobre as normas tem focado em se, e em que medida, outras espécies têm capacidades de sustentar cultura cumulativa (DEAN *et al*., 2014). Razões plausíveis são oferecidas em favor da ideia de que os recursos psicológicos básicos estão presentes não apenas em alguns grandes primatas, mas também em pássaros canoros (WHITEN, 2019) e em baleias e golfinhos (WHITEHEAD; RENDELL, 2015). Tem havido também interesse sobre aspectos da cognição moral: trabalhos bem iluminantes vêm explorando as continuidades entre humanos e outros animais (DE WALL, 2006; ANDREWS; MONSÓ, 2010). Embora não a aborde diretamente, muito disso é relevante para a questão de se a capacidade psicológica dedicada às *normas* é distintivamente humana ou quais de seus mecanismos componentes podem estar presentes, em forma rudimentar, em outros animais. Alguns têm sugerido que a propensão a punir, especialmente a tendência à sanção por violação de normas por parte de terceiros, não é encontrada em outras espécies (RIEDL *et al.* 2012; PROOIJEN, 2018; SUCHAK *et al.*, 2016). Outros apontam para a excepcional habilidade dos seres humanos para cooperar e para a dominância ecológica resultante, sugerindo que isso fornece evidência indireta a favor da singularidade das nossas capacidades para a transmissão cultural de normas (HENRICH, J., 2015; BOYD, 2017). Nem todos se convenceram, contudo, e alguns argumentam que animais como os elefantes (ROSS, 2019) e chimpanzés (VON ROHR *et al.*, 2011) exibem comportamentos socialmente sofisticados que são mais bem explicados pela presença de precursores psicológicos dos componentes básicos do sistema de normas humano. Progressos preliminares importantes têm sido feitos nesse agregado de questões relativas à normatividade não humana (VINCENT; RING; ANDREWS, 2018; ANDREWS, 2020; FITZPATRICK, 2020), mas muito do trabalho conceitual e empírico ainda está por ser feito.

3. Cognição Normativa e Moralidade

Normas são relevantes para áreas de pesquisa que incluem a filosofia, as humanidades e as ciências comportamentais, e os tipos de abordagens cognitivo-comportamentais da psicologia das normas descritas aqui têm o potencial de informar e enriquecer muitas dessas áreas. As implicações mais imediatas supostamente recairiam sobre o domínio da teoria moral. No entanto, a relação das normas e da psicologia normativa com a moralidade e a psicologia moral não é tão direta, e ela própria é objeto de debate (MACHERY, 2012). A busca pela delimitação das fronteiras do domínio moral e por distinguir normas genuinamente morais das regras não morais tem uma longa história, mas ainda está por produzir uma perspectiva que seja amplamente aceita (STICH, 2018). Por exemplo, alguns pesquisadores argumentam que há diferenças psicológicas proximais que podem ser usadas para distinguir um conjunto de regras morais de outras (convenções, regras de etiqueta, regras pragmáticas). De acordo como uma perspectiva importante enraizada na psicologia comportamental, regras morais são delimitadas pelo fato de os indivíduos julgarem que sua aplicação é geral e não apenas local, que sua aplicação é independente da importância de qualquer figura de autoridade, e que elas regem questões relativas a dano, bem-estar, justiça e direitos (TURIEL, 1983; NUCCI, 2001). Alguns têm se inspirado na tradição sentimentalista da teoria moral para oferecer a essa abordagem uma base, explicando assim os aspectos postulados como distintivos das regras morais por meio da sua conexão com emoções como raiva e repugnância (NICHOLS, 2004; cf. HAIDT, 2001). Outros têm contestado a caracterização inicial do que são as normas morais, mobilizando argumentos e evidências de que elas não são psicologicamente universais, mas antes restritas a certas culturas (KELLY *et al.,* 2007; KELLY; STICH, 2007, BERNIŪNAS; DRANSEIKA; SOUSA, 2016; BERNIŪNAS; SILIUS; DRANSEIKA, 2020; KUMAR, 2015; HEATH, 2017).

Mudar o foco das considerações proximais para as distais não parece acrescentar muita clareza. Embora alguns teóricos sustentem que nossa espécie possui um sistema psicológico evoluído dedicado especificamente à moralidade (JOYCE, 2007; MAMELI, 2013; STANFORD, 2018; KITCHER, 2011), outros permanecem céticos. Eles argumentam que as evidências apoiam melhor a perspectiva de que os humanos têm um sistema psicológico evoluído dedicado a normas de maneira geral, mas que não há nada a respeito dos mecanismos que

lhes dão base, das pressões adaptativas que as selecionam, ou das normas que ele pode vir a conter, que ofereceria uma distinção entre normas morais e normas não morais (MACHERY; MALLON, 2010; DAVIS; KELLY, 2018; STICH, 2018). Nessa perspectiva, o sistema humano de normas evoluiu para ser capaz de lidar, sendo ainda capaz de adquirir e internalizar, uma vasta gama de normas, incluindo normas epistêmicas, normas linguísticas, normas de vestimenta, normas religiosas, normas de etiqueta e normas que poderiam ser classificadas por um ocidental contemporâneo como morais. De fato, entende-se que essa proposta apoia uma perspectiva historicista da moralidade, segundo a qual a prática de distinguir algum subconjunto de normas e juízos normativos como morais, que possuem assim um *status* especial ou autoridade, é culturalmente localizado e uma invenção histórica relativamente recente (MACHERY, 2018).

Isso dito, vários debates que têm como objeto tópicos em grande parte morais já começaram a recorrer a abordagens empiricamente informadas da psicologia das normas. Por exemplo, discussões filosóficas das questões morais levantadas pelos vieses sociais implícitos assumiu a forma dos debates respeitados e tradicionais que ocorrem entre individualistas e estruturalistas (BEEGHLY; HOLROYD, 2020; *vide* BROWNSTEIN, 2019)[66] . Uma questão crucial tem sido a de determinar se comportamentos são mais bem explicados e injustiças mais bem tratadas se o foco de atenção for, por um lado, agentes individuais, seus vieses e outras características psicológicas ou, por outro lado, aspectos das instituições ou estruturas sociais habitadas por esses agentes (HASLANGER, 2015). Isso tem inspirado tentativas de desenvolver uma explicação interacionista capaz de combinar as virtudes de ambas as abordagens (MADVA, 2016; SOON, 2020). Várias dessas tentativas colocam as normas em lugar de destaque na análise (AYALA-LÓPEZ, 2018) e se baseiam em pesquisas empíricas a respeito da psicologia normativa a fim de mostrar como as normas funcionam como uma cartilagem que vincula indivíduos e estruturas sociais brandas [*soft social structures*] (DAVIDSON; KELLY, 2020).

Tem-se pensado também que a variação cultural nas normas e as persistentes disputas sobre certo e errado acarretam implicações para a metaética. O "argumento da discordância" (LOEB, 1998) sustenta que se a disputa sobre a permissibilidade de alguma atividade ou prática persistir mesmo após erros de raciocínio e discordâncias

---

[66] N.T.: A primeira edição do verbete de Brownstein é de 2015.

factuais não morais tiverem sido resolvidos, essa discordância intratável ofereceria uma razão contrária ao realismo moral (MACKIE, 1977). Estabelecer empiricamente a existência de discordância persistente é difícil (DORIS; PLAKIAS, 2008), mas o caráter do sistema de normas e sua influência sobre o juízo talvez digam se ele é provável ou não. Considere dois indivíduos de culturas diferentes. Ambos internalizaram famílias diferentes de normas (individualista e coletivista, baseada na honra e baseada na vergonha, divindade e autonomia, rigidez e flexibilidade, etc.). Tais indivíduos provavelmente irão discordar sobre a permissibilidade de uma gama de atividades e práticas, tais como o que conta como uma divisão justa de recursos, se as pessoas podem escolher com quem se casam, ou qual é e qual não é o modo apropriado de responder a um insulto. Essa discordância pode muito bem perdurar mesmo em face da concordância sobre aspectos não morais do problema e mesmo se nenhum dos lados da disputa estiver sendo parcial ou cometendo erros de raciocínio. Essa discordância persistente pode ser explicada por diferenças nos respectivos sistemas de normas desses indivíduos e pelas diferentes normas que cada um internalizou a partir de sua cultura (MACHERY *et al.*, 2005). Detalhes empíricos dos princípios operacionais da cognição normativa – especialmente o de saber se e em que medida ela é informacionalmente encapsulada, cognitivamente impenetrável ou, de algum modo distinto, recalcitrante e insensível a outros processos psicológicos – podem ajudar a avaliar a plausibilidade desse argumento.

Um último conjunto de debates para o qual os detalhes da psicologia das normas têm se tornado cada vez mais relevantes é relativo à natureza e explicação do progresso moral. É comum entender que o progresso moral recente se desenvolveu em grande parte na forma de expansões do círculo moral, na disseminação de normas inclusivas e na desmoralização de normas inválidas (SINGER, 1981; BUCHANAN; POWELL, 2018; SAUER, 2019). Muita atenção tem sido dada à compreensão das mudanças na distribuição de normas que resultam do **raciocínio** sobre as próprias normas (CAMPBELL; KUMAR, 2012), mas passos importantes em direção ao progresso moral podem também ocorrer como resultado de processos míopes (embora não totalmente cegos) de evolução cultural (KLING, 2016; BROWNSTEIN; KELLY, 2019). Uma compreensão empírica mais detalhada da relação das normas internalizadas com itens como racionalização, raciocínio crítico e argumentação explícita (SUMMERS, 2017; MERCIER; SPERBER, 2017), juntamente com uma perspectiva mais clara dos outros fatores psicológicos e sociais que influenciam as normas e as dinâmicas de sua transmissão, irá ajudar a iluminar ainda mais esses debates filosóficos importantes.

Suplemento: descrição estendida da figura

O diagrama tem duas metades. A primeira metade é rotulada como "Mecanismo de Aquisição" e a segunda como "Mecanismo de Execução". Mecanismo de Aquisição tem uma caixa dividida por uma linha pontilhada. A primeira metade da caixa contém o texto "infira conteúdos de regras normativas" e a segunda metade contém o texto "identifique comportamento que implica norma". Abaixo dessa caixa há o texto "Sinais Proximais no Ambiente" com uma seta apontando para o texto da caixa. Uma seta também aponta a partir da caixa para a caixa maior em Mecanismo de Execução.

Mecanismo de Execução consiste em uma caixa grande contendo o texto:

Base de dados de normas
r1 …
r2 …
r3 …
…&
r_n e uma linha pontilhada
Capacidade de raciocínio relacionada a normas

Duas caixas pequenas são adjacentes a essa caixa grande. Uma próxima a seu início tem o texto "motivação de obediência" e, a outra, próxima ao fim, tem o texto "motivação punitiva".

Referência bibliográfica

ANDREWS, K. Pluralistic Folk Psychology in Humans and Other Ape. *In:* KIVERSTEIN, J. (ed.). **The Routledge Handbook of Philosophy of the Social Mind.** New York: Routledge Press, 2017, p. 117–138.

KIVERSTEIN, J. (ed.). Naïve Normativity: The Social Foundation of Moral Cognition. **Journal of the American Philosophical Association**, v. 6, n.1, p. 36–56, 2020. doi:10.1017/apa.2019.30.

ANDREWS, K.; MONSÓ, S. Animal Moral Psychologies. *In:* VARGAS, M.; DORIS, J. M. (ed.). (no prelo). **The Oxford Handbook of Moral Psychology**.

Oxford: Oxford University Press, 2010.

ARIEW, A. Ernst Mayr's "Ultimate/Proximate" Distinction Reconsidered and Reconstructed. **Biology & Philosophy**, v. 18, n. 4, p. 553–565, 2003,. doi:10.1023/A:1025565119032.

AYALA-LÓPEZ, S. A Structural Explanation of Injustice in Conversations: It's about Norms. **Pacific Philosophical Quarterly**, v. 99, n. 4, p. 726–748, 2018. doi:10.1111/papq.12244.

AYARS, A.; NICHOLS, S. Moral Empiricism and the Bias for Act-Based Rules. **Cognition**, v. 167, p. 11–24, 2017. doi:10.1016/j.cognition.2017.01.007

AYARS, A.; NICHOLS, S. Rational Learners and Metaethics: Universalism, Relativism, and Evidence from Consensus. **Mind & Language**, v. 35, n. 1, p. 67–89, 2020. doi:10.1111/mila.12232.

BARKOW, J. H.; COSMIDES, L.; TOOBY, J. **The Adapted Mind**: Evolutionary Psychology and the Generation of Culture. New York: Oxford University Press, 1992.

BARRETT, H. C. *et al.* Small-Scale Societies Exhibit Fundamental Variation in the Role of Intentions in Moral Judgment. **Proceedings of the National Academy of Sciences**, v. 113, n. 17, p. 4688–4693, 2016. doi:10.1073/pnas.1522070113.

BAUMGARTNER, T. *et al.* The Neural Circuitry of a Broken Promise. **Neuron**, v. 64, n. 5, p. 756–770, 2009. doi:10.1016/j.neuron.2009.11.017.

BEDNAR, J. *et al.* Emergent Cultural Signatures and Persistent Diversity: A Model of Conformity and Consistency. **Rationality and Society,** v. 22, n. 4, p. 407–444, 2010. doi:10.1177/1043463110374501

BEEGHLY, E.; HOLROYD, J. Bias in Context: An Introduction to the Symposium. **Journal of Applied Philosophy**, v. 37, n. 2, p. 163–168, 2020. doi:10.1111/japp.12424.

BELLER, S. Deontic Reasoning Reviewed: Psychological Questions, Empirical Findings, and Current Theories. **Cognitive Processing**, v. 11, n. 2, p. 123–132, 2010. doi:10.1007/s10339-009-0265-z

BERNIŪNAS, R.; DRANSEIKA, V.; SOUSA, P. Are There Different Moral Domains? Evidence from Mongolia: Moral Domains. **Asian Journal of Social Psychology**, v. 19, n. 3, p. 275–282, 2016. doi:10.1111/ajsp.12133

BERNIŪNAS, R.; SILIUS, V.; DRANSEIKA, V. Beyond the Moral Domain: The Normative Sense Among the Chinese. **Psichologija**, v. 60, p. 86–105, 2020. doi:10.15388/Psichol.2019.11.

BICCHIERI, C. **The Grammar of Society**: The Nature and Dynamics of Social Norms. Cambridge: Cambridge University Press, 2006. doi:10.1017/CBO9780511616037.

BICCHIERI, C. **Norms in the Wild**: How to Diagnose, Measure, and Change Social Norms. Oxford: Oxford University Press, 2016. doi:10.1093/acprof:oso/9780190622046.001.0001.

BICCHIERI, C.; MULDOON R.; SONTUOSO, A. Social Norms. *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/social-norms/.

BONE, J.; SILVA, A. S.; RAIHANI, N. J. Defectors, Not Norm Violators, Are Punished by Third-Parties. **Biology Letters,** v. 10, n. 7, 20140388, 2014. doi:10.1098/rsbl.2014.0388.

BONE, J. E.; MCAULIFFE, K.; RAIHANI, N. J. Exploring the Motivations for Punishment: Framing and Country-Level Effects. **PLOS ONE**, v. 11, n. 8, e0159769, 2016. doi:10.1371/journal.pone.0159769.

BOYD, R. **A Different Kind of Animal**: How Culture Transformed Our Species. Princeton: Princeton University Press, 2017. (The University Center for Human Values Series).

BOYD, R.; RICHERSON, P. J. Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups. **Ethology and Sociobiology**, v. 13. n. 3, p. 171–195, 1992. doi:10.1016/0162-3095(92)90032-Y

BOYD, R.; RICHERSON, P. J. Norms and Bounded Rationality. *In:* GIGERENZER, G.; SELTEN, R. (ed.). **Bounded Rationality**: The Adaptive Toolbox, Cambridge, MA: The MIT Press, 2001, p. 181–296.

GIGERENZER, G.; SELTEN, R. (ed.). **The Origin and Evolution of Cultures.** New York: Oxford University Press, 2005a.

GIGERENZER, G.; SELTEN, R. (ed.). Solving the Puzzle of Human Cooperation. *In:* LEVINSON, S. C.; JAISSON, P. **Evolution and Culture**. Cambridge, MA: MIT Press, 2005b, p. 105–132.

GIGERENZER, G.; SELTEN, R. (ed.). Gene-Culture Coevolution and the Evolution of Social Institutions, *In:* ENGEL, C.; SINGER, W. (eds). **Better than Conscious?** Decision Making, the Human Mind, and Implications for Institutions. Cambridge, MA: MIT Press, 2008, p. 305–324.

BOYD, R.; RICHERSON, P.J.; HENRICH, J. The Cultural Niche: Why Social Learning is Essential for Human Adaptation. **Proceedings of the National Academy**

of Sciences, v. 108, Supplement 2, p. 10918–10925, 2011. doi:10.1073/pnas.1100290108.

BRENNAN, G. *et al*, **Explaining Norms**. Oxford: Oxford University Press, 2013. doi:10.1093/acprof:oso/9780199654680.001.0001.

BROWN, D. **Human Universals.** New York: McGraw-Hill, 1991.

BROWNSTEIN, M. Implicit Bias *In:* ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy.** Fall Edition. Stanford, CA: The Metaphysics Research Lab, 2019. Disponível em: https://plato.stanford.edu/archives/fall2019/entries/implicit-bias/ (revisão da edição de 2015).

BROWNSTEIN, M.; KELLY, D. **Review of The Evolution of Moral Progress**: A Biocultural Theory by Allen Powell and James Buchanan. The British Journal for the Philosophy of Science: Review of Books, 2018. Disponível em: .

BROWNSTEIN, M.; SAUL, J. (eds.). **Implicit Bias and Philosophy**. Oxford: Oxford University Press, 2016. (2 volumes).

BUCHANAN, A.; POWELL, R. **The Evolution of Moral Progress**: A Biocultural Account. Oxford: Oxford University Press, 2018. doi:10.1093/oso/9780190868413.001.0001.

BURNHAM, T. C.; JOHNSON, D. D. P. The Biological and Evolutionary Logic of Human Cooperation. **Analyse & Kritik**, v. 27, n. 1, p. 113–135, 2005. doi:10.1515/auk-2005-0107.

CAMPBELL, R.; KUMAR, V. Moral Reasoning on the Ground. **Ethics**, v. 122, n. 2, p. 273–312, 2012. doi:10.1086/663980.

CARRUTHERS, P. **The Architecture of the Mind**. Oxford: Oxford University Press, 2006. doi:10.1093/acprof:oso/9780199207077.001.0001.

CARRUTHERS, P. **The Opacity of Mind**: An Integrative Theory of Self-Knowledge. Oxford: Oxford University Press, 2011. doi:10.1093/acprof:oso/9780199596195.001.0001.

CHENG, J. T. *et al.* Two Ways to the Top: Evidence That Dominance and Prestige Are Distinct yet Viable Avenues to Social Rank and Influence. **Journal of Personality and Social Psychology**, v. 104, n. 1, p. 103–125, 2013. doi:10.1037/a0030398.

CHERNYAK, N.; SOBEL, D. M. "But He Didn't Mean to Do It": Preschoolers Correct Punishments Imposed on Accidental Transgressors. **Cognitive Development**, v. 39, p. 13–20, 2016. doi:10.1016/j.cogdev.2016.03.002

CHUDEK, M.; HENRICH, J. Culture–Gene Coevolution, Norm-Psychology and the Emergence of Human Prosociality. **Trends in Cognitive Sciences**, v. 15,

n. 5, p. 218–226, 2011. doi:10.1016/j.tics.2011.03.003

CHUDEK, M.; ZHAO, W.; HENRICH, J. Culture-Gene Coevolution, Large-Scale Cooperation, and the Shaping of Human Social Psychology. *In:* STERELNY, K. *et al.* (eds.). Cooperation and Its Evolution, Cambridge, MA: The MIT Press, 2013, p. 425–458.

CLAY, Z.; TENNIE, C. Is Overimitation a Uniquely Human Phenomenon? Insights From Human Children as Compared to Bonobos. **Child Development,** v. 89, n. 5, p. 1535–1544, 2018. doi:10.1111/cdev.12857

CONTE, R.; ANDRIGHETTO, G.; CAMPENNÌ, M. (eds.). **Minding Norms**: Mechanisms and Dynamics of Social Order in Agent Societies. Oxford: Oxford University Press, 2013.(Oxford Series on Cognitive Models and Architectures) doi:10.1093/acprof:oso/9780199812677.001.0001

COSMIDES, L.; TOOBY, J. Cognitive Adaptations for Social Exchange. *In:* BARKOW, J. H.; COSMIDES, L.; TOOBY, J. **The Adapted Mind**: Evolutionary Psychology and the Generation of Culture. New York: Oxford University Press, 1992, p. 151–192.

CUMMINS, D. D. Evidence for the Innateness of Deontic Reasoning. **Mind & Language**, v. 11, n. 2, p. 160–190, 1996. doi:10.1111/j.1468-0017.1996.tb00039.x

CUMMINS, R. "How does it work?" vs. "What are the laws?": Two conceptions of psychological explanation. *In:* KEIL, F. C.; WILSON, R. W. (eds). **Explanation and Cognition.** Cambridge, MA: The MIT Press, 2000, p. 117–145.

CURTIN, C. M. *et al.* Kinship Intensity and the Use of Mental States in Moral Judgment across Societies. **Evolution and Human Behavior**, 2020. doi:10.1016/j.evolhumbehav.2020.07.002.

CUSHMAN, F.; YOUNG, L.; GREENE, J. D. Multi-System Moral Psychology. *In:* DORIS, J. M. et al. **The Moral Psychology Handbook**. New York: Oxford University Press, 2010, p. 47–71.

DAVIS, T.; HENNES, E. P.; RAYMOND, L. Cultural Evolution of Normative Motivations for Sustainable Behaviour. **Nature Sustainability**, v. 1, n. 5, p. 218–224, 2018. doi:10.1038/s41893-018-0061-9.

DAVIS, T.; KELLY, D. Norms, Not Moral Norms: The Boundaries of Morality Do Not Matter. **Behavioral and Brain Sciences**, v. 41, n. e101, Commentary on Stanford, 2018. doi:10.1017/S0140525X18000067.

DAVIDSON, L. J.; KELLY, D. Minding the Gap: Bias, Soft Structures, and the Double Life of Social Norms. **Journal of Applied Philosophy**, v. 37, n. 2, p. 190–

210, 2020. doi:10.1111/japp.12351.

DEAN, L. G. *et al.* Human Cumulative Culture: A Comparative Perspective: Human Cumulative Culture. **Biological Reviews**, v. 89, n. 2, p. 284–301, 2014. doi:10.1111/brv.12053.

DENNETT, D. **Brainstorms**. Montgomery, VT: Bradford Books, 1978.

DEVINE, P. G. *et al.* Long-Term Reduction in Implicit Race Bias: a Prejudice Habit-Breaking Intervention. **Journal of Experimental Social Psychology**, v. 48, n. 6, p. 1267–1278, 2012. doi:10.1016/j.jesp.2012.06.003.

DORIS, J. M. et al. **The Moral Psychology Handbook**. New York: Oxford University Press, 2010. doi:10.1093/acprof:oso/9780199582143.001.0001.

DORIS, J. M.; PLAKIAS, A. How to Argue About Disagreement: Evaluative Diversity and Moral Realism. *In:* SINNOTT-ARMSTRONG, W. (ed.) **Moral Psychology.** Cambridge, MA: MIT Press, 2008, v. 2. p. 303–332.

DWYER, S.; HUEBNER, B.; HAUSER, M. D. The Linguistic Analogy: Motivations, Results, and Speculations. **Topics in Cognitive Science**, v. 2, n. 3, p. 486–510, 2010. doi:10.1111/j.1756-8765.2009.01064.x

ELSTER, J. Rationality and Social Norms. **European Journal of Sociology**, v. 32, n. 1, p. 109–129, 1991. doi:10.1017/S0003975600006172.

FEHR, E.; CAMERER, C. F. Social Neuroeconomics: The Neural Circuitry of Social Preferences. **Trends in Cognitive Sciences**, v. 11, n. 10, p. 419–427, 2007. doi:10.1016/j.tics.2007.09.002.

FEHR, E.; FALK, A. Psychological Foundations of Incentives. **European Economic Review**, v. 46, n. 4–5, p. 687–724, 2002. doi:10.1016/S0014-2921(01)00208-2.

FEHR, E.; FISCHBACHER, U. Third-Party Punishment and Social Norms. **Evolution and Human Behavior**, v. 25, n. 2, p. 63–87, 2004. doi:10.1016/S1090-5138(04)00005-4.

FEHR, E.; GÄCHTER, S. Altruistic Punishment in Humans. **Nature**, v. 415, n. 6868, p. 137–140, 2002. doi:10.1038/415137a.

FESSLER, D. M. T.; MACHERY, E. Culture and Cognition. *In:* MARGOLIS, E.; SAMUELS, R.; STICH, S. P. (eds). **The Oxford Handbook of Philosophy of Cognitive Science.** Oxford: Oxford University Press, 2012, p. 503–527.

FITZPATRICK, S. Chimpanzee Normativity: Evidence and Objections. **Biology & Philosophy**, v. 35, n. 45, 2020. doi:10.1007/s10539-020-09763-1.

FODOR, J. A. **Psychological Explanation**: An Introduction to the Philosophy of Psychology. New York: Random House, 1968.

FODOR, J. A. **Modularity of Mind**. Cambridge, MA: MIT Press, 1983.

FOGARTY, L.; CREANZA, N. The Niche Construction of Cultural Complexity: Interactions between Innovations, Population Size and the Environment. Philosophical Transactions of the Royal Society B: **Biological Sciences**, v. 372, n. 1735, 2017. doi:10.1098/rstb.2016.0428.

FRANKISH, K. Dual-Process and Dual-System Theories of Reasoning: Dual-Process and Dual-System Theories of Reasoning. **Philosophy Compass**, v. 5, n. 10, p. 914–926, 2010. doi:10.1111/j.1747-9991.2010.00330.x.

GAUS, G. The Open Society as a Rule-Based Order. **Erasmus Journal for Philosophy and Economics**, v. 9, n. 2, p. 1–13, 2016. doi:10.23941/ejpe.v9i2.225.

GAUS, G.; NICHOLS, S, Moral Learning in the Open Society: The Theory and Practice of Natural Liberty. **Social Philosophy and Policy**, v. 34, n. 1, p. 79–101, 2017. doi:10.1017/S0265052517000048.

GAVRILETS, S.; RICHERSON, P. J. Collective Action and the Evolution of Social Norm Internalization. **Proceedings of the National Academy of Sciences**, v. 114, n. 23, p. 6068–6073, 2017. doi:10.1073/pnas.1703857114.

GELFAND, M. J. **Rule Makers, Rule Breakers**: How Tight and Loose Cultures Wire Our World. New York: Scribner, 2018.

GELFAND, M. J.; HARRINGTON, J. R.; JACKSON, J. C. The Strength of Social Norms Across Human Groups. **Perspectives on Psychological Science**, v. 12, n. 5, p. 800–809, 2017. doi:10.1177/1745691617708631.

GELFAND, M. J.; NISHII, L. H.; RAVER, J. L. On the Nature and Importance of Cultural Tightness-Looseness. **Journal of Applied Psychology**, v. 91, n. 6, p. 1225–1244, 2006. doi:10.1037/0021-9010.91.6.1225.

GELFAND, M. *et al.* Differences between Tight and Loose Cultures: a 33-Nation Study. **Science**, v. 332, n. 6033, p. 1100–1104, 2011. doi:10.1126/science.1197754.

GIGERENZER, G.; SELTEN, R. (eds.). **Bounded Rationality**: The Adaptive Toolbox, Cambridge, MA: The MIT Press, 2001.

GIGERENZER, G. *et al.* **Simple Heuristics that Make Us Smart.** New York: Oxford University Press, 2000.

GINTIS, H. *et al.* (eds.). **Moral Sentiments and Material Interests**: The Foundations of Cooperation in Economic Life. Cambridge, MA: MIT Press, 2005. (Economic Learning and Social Evolution 6).

GÖCKERITZ, S.; SCHMIDT, M. F. H.; TOMASELLO, M. Young Children's Creation and Transmission of Social Norms. **Cognitive Development**, v. 30, p. 81–

95, 2014. doi:10.1016/j.cogdev.2014.01.003.

GOWLETT, J. A.J. The Early Settlement of Northern Europe: Fire History in the Context of Climate Change and the Social Brain. **Comptes Rendus Palevol**, v. 5, n. 1–2, p. 299–310, 2006. doi:10.1016/j.crpv.2005.10.008.

GRAHAM, J. *et al.* Moral Foundations Theory. *In:* GRAHAM, J. *et al.* **Advances in Experimental Social Psychology.** San Diego, CA: Academic Press, 2013, p. 47:55–130. doi:10.1016/B978-0-12-407236-7.00002-4.

GRIFFITHS, P. E. Ethology, Sociobiology, and Evolutionary Psychology. *In:* SAHOTRA S.; PLUTYNSKI, A. (eds.). **A Companion to the Philosophy of Biology**. Oxford, UK: Blackwell Publishing Ltd, 2007, p. 393–414. doi:10.1002/9780470696590.ch21.

HAGEN, E. H.; HAMMERSTEIN, P. Game Theory and Human Evolution: A Critique of Some Recent Interpretations of Experimental Games. **Theoretical Population Biology**, v. 69, n. 3, p. 339–348, 2006. doi:10.1016/j.tpb.2005.09.005.

HAGGER, M. S.; RENTZELAS, P.; CHATZISARANTIS, N. L. D. Effects of Individualist and Collectivist Group Norms and Choice on Intrinsic Motivation. **Motivation and Emotion**, v. 38, n. 2, p. 215–223, 2014. doi:10.1007/s11031-013-9373-2.

HAIDT, J. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. **Psychological Review**, v. 108, n. 4, p. 814–834, 2001. doi:10.1037/0033-295X.108.4.814.

HAIDT, J. **The Righteous Mind**: Why Good People Are Divided by Politics and Religion. New York: Pantheon Books, 2012.

HASLANGER, S. Distinguished Lecture: Social Structure, Narrative and Explanation. **Canadian Journal of Philosophy**, v. 45, n. 1, p. 1–15, 2015. doi:10.1080/00455091.2015.1019176.

HAUN, D. B. M., REKERS, Y.; TOMASELLO, M. Children Conform to the Behavior of Peers; Other Great Apes Stick With What They Know. **Psychological Science**, v. 25, n. 12, p. 2160–2167, 2014. doi:10.1177/0956797614553235.

HAUSER, M. D.; YOUNG, L.; CUSHMAN, F. Reviving Rawls' linguistic analogy, *In:* SINNOTT-ARMSTRONG, W. (ed.) **Moral Psychology.** Cambridge, MA: MIT Press, 2008, v. 2. p. 107–144.

HAWKINS, R. X. D.; GOODMAN, N. D.; GOLDSTONE, R. L. The Emergence of Social Norms and Conventions. **Trends in Cognitive Sciences**, v. 23, n. 2, p. 158–169, 2019. doi:10.1016/j.tics.2018.11.003.

HEATH, J. Morality, Convention and Conventional Morality. **Philosophical Explorations**, v. 20, n. 3, p. 276–293, 2017. doi:10.1080/13869795.2017.1362030.

HENRICH, J. **The Secret of Our Success**: How Culture Is Driving
Human Evolution, Domesticating Our Species, and Making Us Smarter. Princeton, NJ: Princeton University Press, 2015.

HENRICH, J. (no prelo). **The WEIRDest People in the World**: How the West Became Psychologically Peculiar and Particularly Prosperous. New York: Farrar, Straus and Giroux.

HENRICH, J. *et al.* In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. **American Economic Review**, v. 91, n. 2, p. 73–78, 2001. doi:10.1257/aer.91.2.73.

HENRICH, J. *et al.* (eds). **Foundations of Human Sociality**: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. Oxford: Oxford University Press, 2004. doi:10.1093/0199262055.001.0001.

HENRICH, J. *et al.* 'Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. **Behavioral and Brain Sciences**, v. 28, n. 6, p. 795–815, 2005. doi:10.1017/S0140525X05000142.

HENRICH, J.; GIL-WHITE, F. J. The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission. **Evolution and Human Behavior**, v. 22, n. 3, p. 165–196, 2001. doi:10.1016/S1090-5138(00)00071-4.

HENRICH, J.; HEINE, S. J.; NORENZAYAN, A. The Weirdest People in the World? **Behavioral and Brain Sciences**, v. 33, n. 2–3, p. 61–83, 2010. doi:10.1017/S0140525X0999152X.

HENRICH, J.; MCELREATH, R. The Evolution of Cultural Evolution. **Evolutionary Anthropology**: Issues, News, and Reviews, v. 12, n. 3, p. 123–135, 2003. doi:10.1002/evan.10110.

HENRICH, J. *et al.* Costly Punishment Across Human Societies. **Science**, v. 312, n. 5781, p. 1767–1770, 2006. doi:10.1126/science.1127333.

HENRICH, N.; HENRICH, J. **Why Humans Cooperate**: A Cultural and Evolutionary Explanation. Oxford: Oxford University Press, 2007.

HERRMANN, B.; THÖNI, C.; GÄCHTER, S. Antisocial Punishment Across Societies. **Science**, v. 319, n. 5868, p. 1362–1367, 2008. doi:10.1126/science.1153808.

HEYES, C. M. **Cognitive Gadgets**: The Cultural Evolution of Thinking. Cambridge, MA: Harvard University Press, 2018.

HOEHL, S. *et al.* 'Over-Imitation': A Review and Appraisal of a Decade of Research. **Developmental Review**, n. 51, p. 90–108, 2019. doi:10.1016/j.dr.2018.12.002.

HOFSTEDE, G. **Culture's Consequences**: International Differences in Work Related Values. Beverly Hills, CA: SAGE, 1980.

HOFSTEDE, G. **Culture's Consequences**: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Thousand Oaks, CA: SAGE, 2001.

HOUSE, B. R. *et al.* Universal Norm Psychology Leads to Societal Diversity in Prosocial Behaviour and Development. **Nature Human Behaviour**, v. 4, n. 1, p. 36–44, 2020. doi:10.1038/s41562-019-0734-z.

HOUSE, B. R. *et al.* Ontogeny of Prosocial Behavior across Diverse Societies. **Proceedings of the National Academy of Sciences**, v. 110, n. 36, p. 14586–14591. doi:10.1073/pnas.1221217110.

INGLEHART, R. **Modernization and Postmodernization**: Cultural, Economic, and Political Change in 43 Societies. Princeton, NJ: Princeton University Press, 1997.

JENSEN, K.; CALL, J.; TOMASELLO, M. Chimpanzees Are Rational Maximizers in an Ultimatum Game. **Science**, v. 318, n. 5847, p. 107–109, 2007. doi:10.1126/science.1145850.

JOSEPHS, M. *et al.* Children Protest Moral and Conventional Violations More When They Believe Actions Are Freely Chosen. **Journal of Experimental Child Psychology**, n. 141, p. 247–255, 2016. doi:10.1016/j.jecp.2015.08.002.

JOYCE, R. **The Evolution of Morality.** Cambridge, MA: MIT Press, 2007.

KAHNEMAN, D. **Thinking Fast and Slow.** New York: Farrar, Straus and Giroux, 2011.

KELLY, D. R. **Yuck!** The Nature and Moral Significance of Disgust. Cambridge, MA: The MIT Press, 2011.

KELLY, D. R. Internalized Norms and Intrinsic Motivation: Are Normative Motivations Psychologically Primitive? **Emotion Review**, jun., p. 36-45, 2020.

KELLY, D. R. Two Ways to Adopt a Norm: The (Moral?) Psychology of Avowal and Internalization. *In:* VARGAS, M.; DORIS, J. M. (ed.). (no prelo). **The Oxford Handbook of Moral Psychology**. Oxford: Oxford University Press.

KELLY, D.; DAVIS, T. Social Norms and Human Normative Psychology. **Social Philosophy and Policy**, v. 35, n. 1, p. 54–76, 2018. doi:10.1017/S0265052518000122.

KELLY, D. R.; STICH, S. Two Theories of the Cognitive Architecture Underlying Morality. *In:* CARRUTHERS, P.; LAURENCE, S.; STICH, S. (eds). **The Innate Mind**. New York: Oxford University Press, 2007, v. 2, p. 348–366.

KELLY, D. *et al.* Harm, Affect, and the Moral/Conventional Distinction. **Mind & Language**, v. 22, n. 2, p. 117–131, 2007. doi:10.1111/j.1468-0017.2007.00302.x.

KENWARD, B. Over-Imitating Preschoolers Believe Unnecessary Actions Are Normative and Enforce Their Performance by a Third Party. **Journal of Experimental Child Psychology**, v. 112, n. 2, p. 195–207, 2012. doi:10.1016/j.jecp.2012.02.006.

KENWARD, B.; KARLSSON, M.; PERSSON, J. Over-Imitation is Better Explained by Norm Learning than by Distorted Causal Learning. **Proceedings of the Royal Society B**: Biological Sciences, v. 278, n. 1709, p. 1239–1246, 2011. doi:10.1098/rspb.2010.1399.

KEUPP, S.; BEHNE, T.; RAKOCZY, H. Why Do Children Overimitate? Normativity Is Crucial. **Journal of Experimental Child Psychology**, v. 116, n. 2, p. 392–406, 2013. doi:10.1016/j.jecp.2013.07.002.

KITCHER, P. **The Ethical Project**. Cambridge, MA: Harvard University Press, 2011.

KLINE, M. A.; BOYD, R. Population Size Predicts Technological Complexity in Oceania. **Proceedings of the Royal Society B**: Biological Sciences, v. 277, n. 1693, p. 2559–2564, 2010. doi:10.1098/rspb.2010.0452.

KLING, A. Cultural Intelligence. **National Affairs**, n. 27, p. 150–163, 2016.

KRASNOW, M. M., *et al.* Group Cooperation without Group Selection: Modest Punishment Can Recruit Much Cooperation. **PLOS ONE**, v. 10, n. 4, 2015. doi:10.1371/journal.pone.0124561.

KUMAR, V. Moral Judgment as a Natural Kind. **Philosophical Studies**, v. 172, n. 11, p. 2887–2910, 2015. doi:10.1007/s11098-015-0448-7.

LALAND, K. N. **Darwin's Unfinished Symphony**: How Culture Made the Human Mind. Princeton, NJ: Princeton University Press, 2017.

LALAND, K. N.; ODLING-SMEE, J.; FELDMAN, M. W. Cultural Niche Construction and Human Evolution: Niche Construction and Human Evolution. **Journal of Evolutionary Biology**, v. 14, n. 1, p. 22–33, 2001. doi:10.1046/j.1420-9101.2001.00262.x.

LALAND, K. N.; ODLING-SMEE, J.; MYLES, S. How Culture Shaped the Human Genome: Bringing Genetics and the Human Sciences Together. **Nature Reviews Genetics**, v. 11, n. 2, p. 137–148, 2010. doi:10.1038/nrg2734.

LAI, C. K. *et al.* Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions. **Journal of Experimental Psychology**: General, v. 143, n. 4, p. 1765–1785, 2014. doi:10.1037/a0036260

LAMM, E. Forever United: The Coevolution of Language and Normativity. *In:* DOR, D.; KNIGHT, C.; LEWIS, J. (eds). **The Social Origins of Language**: Early Society, Communication and Polymodality. Oxford: Oxford University Press, 2014, p. 267–283.

LESOROGOL, C. K. Bringing Norms in: The Role of Context in Experimental Dictator Games. **Current Anthropology**, v. 48, n. 6, p. 920–926, 2007.

LAURENCE, S.; MARGOLIS, E. The Poverty of the Stimulus Argument. **The British Journal for the Philosophy of Science**, v. 52, n. 2, p. 217–276, 2001. doi:10.1093/bjps/52.2.217.

LEVIN, J. Functionalism. *In:* ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy.** Fall Edition. Stanford, CA: The Metaphysics Research Lab, 2018. Disponível em: https://plato.stanford.edu/archives/fall2018/entries/functionalism/. (revisão da edição de 2004).

LINDSTRÖM, B. *et al.* The Role of a 'Common Is Moral' Heuristic in the Stability and Change of Moral Norms. **Journal of Experimental Psychology**: General, v. 147, n. 2, p. 228–242, 2018. doi:10.1037/xge0000365.

LOEB, D. Moral Realism and the Argument from Disagreement. **Philosophical Studies**, v. 90, n. 3, p. 281–303, 1998. doi:10.1023/A:1004267726440.

LYCAN, W. G. The Continuity of Levels of Nature. *In:* LYCAN, W. G. (ed). **Mind and Cognition**: A Reader. Oxford: Blackwell Publishers, 1990, p. 77–96.

LYONS, D. E.; YOUNG, A. G.; KEIL, F. C. The Hidden Structure of Over imitation. **Proceedings of the National Academy of Sciences**, v. 104, n. 50, p. 19751–19756, 2007. doi:10.1073/pnas.0704452104.

MACHERY, E. Delineating the Moral Domain, Baltic International Yearbook of Cognition. **Logic and Communication**, v. 7 n. 1, 2012. doi:10.4148/biyclc.v7i0.1777.

MACHERY, E. Morality: A Historical Invention. *In:* GRAY, K.; GRAHAM, J. (eds.). **Atlas of Moral Psychology**. New York, NY: Guilford Press, 2018, p. 259–265.

MACHERY, E.; KELLY, D.; STICH, S. P. Moral Realism and Cross-Cultural Normative Diversity. **Behavioral and Brain Sciences**, v. 28, n. 6, p. 830–830, 2005. doi:10.1017/S0140525X05370142.

MACHERY, E.; MALLON, R.Evolution of Morality. *In:* DORIS, J. M. et al. **The Moral Psychology Handbook**. New York: Oxford University Press, 2010, p. 3–46.

MACKIE, J. L. **Ethics**: Inventing Right and Wrong. New York: Pelican Books, 1977.

MADVA, A. A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy. **Ergo**, v. 3, n. 27, p. 701–728, 2016. doi:10.3998/

ergo.12405314.0003.027.

MAMELI, M. Meat Made Us Moral: A Hypothesis on the Nature and Evolution of Moral Judgment. **Biology & Philosophy**, v. 28, n. 6, p. 903–931, 2013. doi:10.1007/s10539-013-9401-3.

MANER, J. K. Dominance and Prestige: A Tale of Two Hierarchies. **Current Directions in Psychological Science**, v. 26, n. 6, p. 526–531, 2017. doi:10.1177/0963721417714323.

MARR, D. **Vision**: A Computational Investigation into the Human Representation and Processing of Visual Information. Cambridge, MA: The MIT Press, 1982.

MARWELL, G.; AMES, R. E. Economists Free Ride, Does Anyone Else? **Journal of Public Economics**, v. 15, n. 3, p. 295–310, 1981. doi:10.1016/0047-2727(81)90013-X.

MATHEW, S.; BOYD, R. Punishment Sustains Large-Scale Cooperation in Prestate Warfare. **Proceedings of the National Academy of Sciences**, v. 108, n. 28, p. 11375–11380, 2011. doi:10.1073/pnas.1105604108.

MATHEW, S.; BOYD, R.; VAN VEELEN, M. Human Cooperation Among Kin and Close Associates May Require Enforcement of Norms by Third Parties. *In:* RICHERSON, P. J.; CHRISTIANSEN, M. H. (ed.). **Cultural Evolution**: Society, Technology, Language, and Religion. Cambridge, MA: MIT Press, 2013, p. 45–60. (Strüngmann Forum Report 12).

MATHEW, S.; PERREAULT, C. Behavioural Variation in 172 Small-Scale Societies Indicates that Social Learning is the Main Mode of Human Adaptation. **Proceedings of the Royal Society B**: Biological Sciences, v. 282, n. 1810, 2015. doi:10.1098/rspb.2015.0061.

MAYR, E. Cause and Effect in Biology. **American Association for the Advancement of Science**, v. 134, n. 3489, p. 1501–1506, 1961.

MCAULIFFE, B. J. *et al.* Individualist and Collectivist Norms: When It's Ok to Go Your Own Way. **European Journal of Social Psychology**, v. 33, n. 1, p. 57–70, 2003. doi:10.1002/ejsp.129.

MCAULIFFE, K.; JORDAN, J. J.; WARNEKEN, F. Costly Third-Party Punishment in Young Children. **Cognition**, n. 134, p. 1–10, 2015. doi:10.1016/j.cognition.2014.08.013.

MCELREATH, R.; BOYD, R.; RICHERSON, P. J. Shared Norms and the Evolution of Ethnic Markers. **Current Anthropology**, v. 44, n. 1, p. 122–130, 2003. doi:10.1086/345689.

MERCIER, H.; SPERBER, D. **The Enigma of Reason**. Cambridge, MA: Harvard University Press, 2017.

MIKHAIL, J. Universal Moral Grammar: Theory, Evidence and the Future. **Trends in Cognitive Sciences**, v. 11, n. 4, p. 143–152, 2007. doi:10.1016/j.tics.2006.12.007.

MIKHAIL, J. **Elements of Moral Cognition**: Rawls'' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment. Cambridge: Cambridge University Press, 2011. doi:10.1017/CBO9780511780578.

MORRIS, M. W. *et al.* Normology: Integrating Insights about Social Norms to Understand Cultural Dynamics. **Organizational Behavior and Human Decision Processes**, n. 129, p. 1–13, 2015. doi:10.1016/j.obhdp.2015.03.001.

MOYA, C.; BOYD, R. Different Selection Pressures Give Rise to Distinct Ethnic Phenomena: a Functionalist Framework with Illustrations from the Peruvian Altiplano. **Human Nature**, v. 26, n. 1, p. 1–27, 2015. doi:10.1007/s12110-015-9224-9.

MUTHUKRISHNA, M.; MORGAN, T. J. H.; HENRICH, J. The When and Who of Social Learning and Conformist Transmission. **Evolution and Human Behavior**, v. 37, n. 1, p. 10–20, 2016. doi:10.1016/j.evolhumbehav.2015.05.004.

NICHOLS, S. **Sentimental Rules**: On the Natural Foundations of Moral Judgement. New York: Oxford University Press, 2004. doi:10.1093/0195169344.001.0001.

NICHOLS, S. **Rational Rules**: Towards a Theory of Moral Learning. Oxford: Oxford University Press, 2021.

NICHOLS, S. *et al.* Rational Learners and Moral Rules: Rational Learners and Moral Rules. **Mind & Language**, v. 31, n. 5 p. 530–554, 2016. doi:10.1111/mila.12119.

NIELSEN, M.; KAPITÁNY, R.; ELKINS, R. The Perpetuation of Ritualistic Actions as Revealed by Young Children's Transmission of Normative Behavior. **Evolution and Human Behavior**, v. 36, n. 3, p. 191–198. doi:10.1016/j.evolhumbehav.2014.11.002.

NISBETT, R. E. **The Geography of Thought**: how Asians and Westerners Think Differently…and why. New York: Free Press, 2004.

NISBETT, R. E.; COHEN, D. Culture of Honor: the Psychology of Violence in the South. Boulder, CO: Westview Press, 1996.

NUCCI, L. P. **Education in the Moral Domain**. Cambridge: Cambridge University Press, 2001. doi:10.1017/CBO9780511605987.

O'GORMAN, R.; WILSON, D. S.; MILLER, R. R. An Evolved Cognitive Bias for Social Norms. **Evolution and Human Behavior**, v. 29, n. 2, p. 71–78,

2008. doi:10.1016/j.evolhumbehav.2007.07.002.

O'NEILL, E. Kinds of Norms, Philosophy Compass, v. 12, n. 5, 2017. doi:10.1111/phc3.12416.

O'NEILL, E.; MACHERY, E. The Normative Sense: What is Universal? What Varies? *In:* ZIMMERMAN, A.; JONES, K.; TIMMONS, M. (ed.). **The Routledge Handbook of Moral Epistemology.** Abingdon, UK: Routledge Press, 2018, cap. 2.

OOSTERBEEK, H.; SLOOF, R.; VAN DE KUILEN, G. Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. **Experimental Economics**, v. 7 n. 2, p. 171–188, 2004. doi:10.1023/B:EXEC.0000026978.14316.74

PEACH, J.; YOSHIDA, E.; ZANNA, M. P. Learning What Most People like: How Implicit Attitudes and Normative Evaluations Are Shaped by Motivation and Culture and Influence Meaningful Behavior. *In:* FORGAS, J. P. *et al.* (eds.). **The Psychology of Attitudes and Attitude Change.** Philadelphia: Psychology Press, 2011, p. 95–108.

PINKER, S. **Enlightenment Now**: The Case for Reason, Science, Humanism, and Progress. New York: Viking Press, 2018.

PRICE, M. E; COSMIDES, L.; TOOBY, J. Punitive Sentiment as an Anti-Free Rider Psychological Device. **Evolution and Human Behavior**, v. 23, n. 3, p. 203–231, 2002. doi:10.1016/S1090-5138(01)00093-9.

PRINZ, J. J. Resisting the Linguistic Analogy: A Commentary on Hauser, Young, and Cushman. *In:* SINNOTT-ARMSTRONG, W. (ed.) **Moral Psychology.** Cambridge, MA: MIT Press, 2008, v. 2. 157–170.

PROOIJEN, J.-W. van. **The Moral Punishment Instinct**. Oxford: Oxford University Press, 2018. doi:10.1093/oso/9780190609979.001.0001.

PUTNAM, H. Brains and Behavior. *In:* BUTLER, R. J. (ed.(. **Analytical Philosophy**. Oxford: Basil Blackwell, p. 325–341. (revisão da edição de 1963).

PUTNAM, H. 1967 [1975], Psychological Predicates, in Art, Mind and Religion, W. H. Capitan and D. D. Merrill (eds), Pittsburgh, PA: University of Pittsburgh Press, 37–48. Reprinted as The Nature of Mental States, in Putnam 1975: 429–440.

PUTNAM, H. Philosophical Papers, Volume 2: Mind, Language, and Reality, Cambridge: Cambridge University Press. doi:10.1017/CBO9780511625251.018.

DE QUERVAIN, D. J.-F. *et al.* The Neural Basis of Altruistic Punishment.

Science, v. 305, n. 5688, p. 1254–1258, 2004. doi:10.1126/science.1100735.

RAKOCZY, H. *et al.* Bigger Knows Better: Young Children Selectively Learn Rule Games from Adults Rather than from Peers. **British Journal of Developmental Psychology**, v. 28, n. 4, p. 785–798, 2010. doi:10.1348/026151009X479178.

RAKOCZY, H.; SCHMIDT, M. F. H. The Early Ontogeny of Social Norms. **Child Development Perspectives**, v. 7, n. 1, p. 17–21, 2013. doi:10.1111/cdep.12010.

RAKOCZY, H.; WARNEKEN, F.; TOMASELLO, M. The Sources of Normativity: Young Children's Awareness of the Normative Structure of Games.
Developmental Psychology, v. 44, n. 3, p. 875–881, 2008. doi:10.1037/0012-1649.44.3.875.

RAMSEY, G. Culture in Humans and Other Animals. **Biology & Philosophy**, v. 28, n. 3, p. 457–479, 2013. doi:10.1007/s10539-012-9347-x.

RICHERSON, P. J. Human Cooperation is a Complex Problem with Many Possible Solutions: Perhaps All of Them Are True! **Cliodynamics**: The Journal of Theoretical and Mathematical History, v. 4, n. 1, p. 139–152, 2013.

RICHERSON, P. *et al.* Cultural Group Selection Plays an Essential Role in Explaining Human Cooperation: A Sketch of the Evidence. **Behavioral and Brain Sciences**, n. 39, e30, 2016. doi:10.1017/S0140525X1400106X.

RICHERSON, P. J.; BOYD, R. The Evolution of Subjective Commitment to Groups: A Tribal Instincts Hypothesis. *In:* NESSE, R. M. (ed.). **The Evolution and the Capacity for Commitment.** New York, NY: Russell Sage, 2001, p. 186–220.

RICHERSON, P. J.; BOYD, R. **Not By Genes Alone**: How Culture Transformed Human Evolution. Chicago: University of Chicago Press, 2005.

RICHERSON, P. J.; BOYD, R. Rethinking Paleoanthropology: a World Queerer Than We Supposed. *In:* HATFIELD, G.; PITTMAN H. (eds.). **Evolution of Mind, Brain, and Culture.** Philadelphia, PA: University of Pennsylvania Museum of Archaeology and Anthropology, 2013, p. 263–302.

RICHERSON, P. J.; BOYD, R.; HENRICH, J. Gene-Culture Coevolution in the Age of Genomics. **Proceedings of the National Academy of Sciences**, n. 107, Supplement 2, p. 8985–8992, 2010. doi:10.1073/pnas.0914631107

RICHERSON, P. J.; HENRICH, J. Tribal Social Instincts and the Cultural Evolution of Institutions to Solve Collective Action Problems. **Cliodynamics**: The Journal of Theoretical and Mathematical History, v. 3, n. 1, p. 38–80, 2012.

RIEDL, K. *et al.* No Third-Party Punishment in Chimpanzees. **Proceedings of the National Academy of Sciences**, v. 109, n. 37, p. 14824–14829, 2012. doi:10.1073/pnas.1203179109.

ROBBINS, P. Modularity of Mind. *In:* ZALTA, E. N. (ed.). **Stanford Encyclopedia of Philosophy.** Winter Edition. Stanford, CA: The Metaphysics Research Lab, 2017. Disponível em: https://plato.stanford.edu/archives/win2017/entries/modularity-mind/. (revisão da edição de 2009).

ROBERTS, S. O.; GELMAN, S. A.; HO, A. K. So It Is, So It Shall Be: Group Regularities License Children's Prescriptive Judgments. **Cognitive Science**, n. 41, Supplement 3, p. 576–600, 2017. doi:10.1111/cogs.12443.

ROBERTS, S. O. *et al.* Children's Descriptive-to-Prescriptive Tendency Replicates (and Varies) Cross-Culturally: Evidence from China. **Journal of Experimental Child Psychology**, n. 165, p. 148–160, 2018. doi:10.1016/j.jecp.2017.03.018.

ROBERTS, S. O.; HO, A. K.; GELMAN, S. A. The Role of Group Norms in Evaluating Uncommon and Negative Behaviors. **Journal of Experimental Psychology**: General, v. 148, n. 2, p. 374–387, 2019. doi:10.1037/xge0000534.

ROEDDER, E.; HARMAN, G. Linguistics and Moral Theory. *In:* DORIS, J. M. et al. **The Moral Psychology Handbook**. New York: Oxford University Press, 2010, p 273 – 296.

VON ROHR, C. R.; BURKART, J. M.; VAN SCHAIK, C. V. Evolutionary Precursors of Social Norms in Chimpanzees: a New Approach. **Biology & Philosophy**, v. 26, n. 1, p. 1–30, 2011. doi:10.1007/s10539-010-9240-4.

ROSS, D. The Evolution of Individualistic Norm. **Baltic International Yearbook of Cognition, Logic and Communication**, n. 7, 2012. doi:10.4148/biyclc.v7i0.1780.

ROSS, D. Consciousness, Language, and the Possibility of Non-Human Personhood: Reflections on Elephants. **Journal of Consciousness Studies**, v. 26, n. 3–4, p. 227–251, 2019.

ROTH, A. E. *et al.* Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. **The American Economic Review**, v. 81, n. 5, p. 1068–1095, 1991.

ROZIN, P. *et al.* The CAD Triad Hypothesis: A Mapping between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity). **Journal of Personality and Social Psychology**, v. 76, n. 4, p. 574–586, 1999. doi:10.1037/0022-3514.76.4.574.

SAMLAND, J. *et al.* The Role of Prescriptive Norms and Knowledge in Children's and Adults' Causal Selection. **Journal of Experimental Psychology**: General, v. 145, n. 2, p. 125–130, 2016. doi:10.1037/xge0000138.

SAMUELS, R. Evolutionary Psychology and the Massive Modularity Hypothesis. **The British Journal for the Philosophy of Science**, v. 49, n. 4, p. 575–602, 1998. doi:10.1093/bjps/49.4.575.

SAUER, H. Butchering Benevolence Moral Progress beyond the Expanding Circle. **Ethical Theory and Moral Practice**, v. 22, n. 1, p. 153–167, 2019. doi:10.1007/s10677-019-09983-9.

SCHMIDT, M. F. H.; RAKOCZY, H.; TOMASELLO, M. Young Children Attribute Normativity to Novel Actions without Pedagogy or Normative Language: Young Children Attribute Normativity. **Developmental Science**, v. 14, n. 3, p. 530–539, 2011. doi:10.1111/j.1467-7687.2010.01000.x.

SCHMIDT, M. F. H. *et al.* Young Children See a Single Action and Infer a Social Norm: Promiscuous Normativity in 3-Year-Olds. **Psychological Science**, v. 27, n. 10, p. 1360–1370, 2016. doi:10.1177/0956797616661182.

SCHMIDT, M. F. H.; TOMASELLO, M. Young Children Enforce Social Norms. **Current Directions in Psychological Science**, v. 21, n. 4, p. 232–236, 2012. doi:10.1177/0963721412448659.

SCHULTZ, P. W. *et al.* The Constructive, Destructive, and Reconstructive Power of Social Norms. **Psychological Science**, v. 18, n. 5, p. 429–434, 2007. doi:10.1111/j.1467-9280.2007.01917.x.

SCHULZ, J. F. *et al.* The Church, Intensive Kinship, and Global Psychological Variation. **Science**, v. 366, n. 6466, eaau5141, 2019. doi:10.1126/science.aau5141.

SHWEDER, R. A. *et al.* The 'Big Three' of Morality (Autonomy, Community, and Divinity), and the 'Big Three' Explanations of Suffering. *In:* BRANDT, A. M.; ROZIN, P. (ed.). **Morality and Health**. New York: Routledge, 1997.

SINGER, P. The Expanding Circle: Ethics, Evolution, and Moral Progress, Princeton, NJ: Princeton University Press, 2011. (revisão da edição de 1981).

SINNOTT-ARMSTRONG, W. (ed.). **Moral Psychology.** Cambridge, MA: MIT Press, 2008, v. 2. (The Cognitive Science of Morality: Intuition and Diversity)

SMETANA, J. G. Understanding of Social Rules. *In:* BENNETT, M. (ed.) **The Development of Social Cognition**: The Child as Psychologist. New York: Guilford Press, 1993, p. 111–141.

SOON, V. Implicit Bias and Social Schema: a Transactive Memory Approach. **Philosophical Studies**, v. 177, n. 7, p. 1857–1877, 2020. doi:10.1007/s11098-019-01288-y.

SPERBER, D.**Explaining Culture**: a Naturalistic Approach. New York: Blackwell Publishers, 1996.

SRIPADA, C. S.; STICH, S. A Framework for the Psychology of Norms. *In:* CARRUTHERS, P.; LAURENCE, S.; STICH, S. (eds.). **The Innate Mind.** New York: Oxford University Press, 2007, v. 2, p. 280–301. (Culture and Cognition).

STANFORD, P. K. The Difference between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation. **Behavioral and Brain Sciences**, n.41, e95, 2018. doi:10.1017/S0140525X17001911.

STERELNY, K. **Thought in a Hostile World**. New York, Blackwell, 2003.

STERELNY, K. **The Evolved Apprentice**: How Evolution Made Humans Unique, Cambridge. MA. The MIT Press, 2012.

STERELNY, K. *et al.* (eds.). Cooperation and Its Evolution, Cambridge, MA: The MIT Press, 2013.

STICH, S. The Quest for the Boundaries of Morality. *In:* ZIMMERMAN, A.; JONES, K.; TIMMONS, M. (ed.). **The Routledge Handbook of Moral Epistemology.** Abingdon, UK: Routledge Press, 2018, cap. 1.

JONES, K.; TIMMONS, M. (ed.). Did Religion Play a Role in the Evolution of Morality? **Religion, Brain & Behavior**, p. 1–11, 2019. doi:10.1080/2153599X.2019.1678511.

SUCHAK, M. *et al.* How Chimpanzees Cooperate in a Competitive World. **Proceedings of the National Academy of Sciences**, v. 113, n. 36, p. 10215–10220, 2016. doi:10.1073/pnas.1611826113.

SUMMERS, J. S. Rationalizing Our Way into Moral Progress. **Ethical Theory and Moral Practice**, v. 20, n. 1, p. 93–104, 2017. doi:10.1007/s10677-016-9750-5

TABIBNIA, G.; SATPUTE, A. B.; LIEBERMAN, M. D. The Sunny Side of Fairness: Preference for Fairness Activates Reward Circuitry (and Disregarding Unfairness Activates Self-Control Circuitry). **Psychological Science**, v. 19, n. 4, p. 339–347, 2008. doi:10.1111/j.1467-9280.2008.02091.x.

TENNIE, C.; CALL, J.; TOMASELLO, M. Ratcheting up the Ratchet: On the Evolution of Cumulative Culture. **Philosophical Transactions of the Royal Society B**: Biological Sciences, v. 364, n. 1528, p. 2405–2415, 2009. doi:10.1098/rstb.2009.0052.

THALER, R. H. **The Winner's Curse**: Paradoxes and Anomalies of Economic Life. New York: Free Press, 1992.

THIEME, H. Lower Palaeolithic Hunting Spears from Germany. **Nature**, v. 385, n. 6619, p. 807–810, 1997. doi:10.1038/385807a0.

TOMASELLO, M. **The Cultural Origins of Human Cognition.** Cambridge, MA: Harvard University Press, 1999.

TOMASELLO, M. **Why We Cooperate**. Cambridge, MA: MIT Press, 2009.

TOMASELLO, M. **A Natural History of Human Morality**. Cambridge, MA: Harvard University Press, 2016.

TOMASELLO, M. **Becoming Human**: A Theory of Ontogeny. Cambridge, MA: Belknap Press, 2019.

TOOBY, J.; COSMIDES, L. The Psychological Foundations of Culture. *In:* BARKOW, J. H.; COSMIDES, L.; TOOBY, J. **The Adapted Mind**: Evolutionary Psychology and the Generation of Culture. New York: Oxford University Press, 1992, p. 19–136.

BARKOW, J. H.; COSMIDES, L.; TOOBY, J. Conceptual Foundations of Evolutionary Psychology. *In:* BUSS, D. M. (ed.).**The Handbook of Evolutionary Psychology.** Hoboken, NJ: John Wiley & Sons, 2005, p. 5–67. doi:10.1002/9780470939376.ch1.

TURCHIN, P. **Historical Dynamics**: Why States Rise and Fall. Princeton, NJ: Princeton University Press, 2018.

TURIEL, E. **The Development of Social Knowledge**. Cambridge: Cambridge University Press, 1983.

TWOREK, C. M.; CIMPIAN, A. Why Do People Tend to Infer 'Ought' From 'Is'? The Role of Biases in Explanation. **Psychological Science**, v. 27, n. 8, p. 1109–1122, 2016. doi:10.1177/0956797616650875.

USKUL, A. K. *et al.* Cultures of Honor. *In:* USKUL, A. K. **Handbook of Cultural Psychology**, 2. ed. New York: The Guilford Press, 2019, p. 793–821.

VAESEN, K. *et al.* Population Size Does Not Explain Past Changes in Cultural Complexity. **Proceedings of the National Academy of Sciences**, v. 113, n. 16, e2241–e2247, 2016. doi:10.1073/pnas.1520288113.

VAISH, A. *et al.* Preschoolers Value Those Who Sanction Non-Cooperators. **Cognition**, n. 153, p. 43–51, 2016. doi:10.1016/j.cognition.2016.04.011.

VARGAS, M.; DORIS, J. M. (eds.). (no prelo). **The Oxford Handbook of Moral Psychology**. Oxford: Oxford University Press.

VINCENT, S.; RING, R.; ANDREWS, K. Normative Practices of Other Animals. *In:* ZIMMERMAN, A.; JONES, K.; TIMMONS, M. (eds.). **The Routledge Handbook of Moral Epistemology.** Abingdon, UK: Routledge Press, 2018. cap. 3, p. 57–83.

DE WAAL, F. **Primates and Philosophers**: How Morality Evolved. Princeton, NJ: Princeton University Press, 2006.

WENDEL, W. B. Mixed Signals: Rational-Choice Theories of Social Norms and the Pragmatics of Explanation. **Washington & Lee Public Law Research Paper**, 01-8, 2001. doi:10.2139/ssrn.278277.

WHITEHEAD, H.; RENDELL, L. **The Cultural Lives of Whales and Dolphins.** Chicago, IL: University of Chicago Press, 2015.

WHITEN, A. Cultural Evolution in Animals. **Annual Review of Ecology, Evolution, and Systematics**, v. 50, n. 1, p. 27–48, 2019. doi:10.1146/annurev-ecolsys-110218-025040.

WHITEN, A. *et al.* Emulation, Imitation, over-Imitation and the Scope of Culture for Child and Chimpanzee. **Philosophical Transactions of the Royal Society B**: Biological Sciences, v. 364, n. 1528, p. 2417–2428, 2009. doi:10.1098/rstb.2009.0069.

YOSHIDA, E. *et al.* Not All Automatic Associations Are Created Equal: How Implicit Normative Evaluations Are Distinct from Implicit Attitudes and Uniquely Predict Meaningful Behavior. **Journal of Experimental Social Psychology**, v. 48, n. 3, p. 694–706, 2012. doi:10.1016/j.jesp.2011.09.013.

ZIMMERMAN, A.; JONES, K.; TIMMONS, M. (eds.). **The Routledge Handbook of Moral Epistemology.** Abingdon, UK: Routledge Press, 2018. doi:10.4324/9781315719696

Outros Recursos da Internet

- MILLGRAM, Elijah, 2019, Bounded Agency, manuscrito não publicado.
- Culture and the Mind, UK Arts & Humanities Research Council.
- Culture, Cognition and Coevolution Lab, Department of Human Evolutionary Biology, Harvard University.
- Tomasello Lab, CHILD Studies group, Department of Psychology and Neuroscience, Duke University.
- Culture Lab, Michele Gelfand, Social Decision and Organizational Sciences (SDOS) group, Psychology Department, University of Maryland/College Park.
- Moral Psychology Research Group.

(VII) Filosofia Moral Experimental*

Autoria: Mark Alfano, Don Loeb, Alexandra Plakias
Tradução: Sagid Salles
Revisão: Roseline Crippa

A filosofia moral experimental surgiu como uma metodologia na última década do século XX, como um ramo da área mais ampla da filosofia experimental (X-Fi). A filosofia moral experimental é o estudo empírico das intuições, juízos e comportamentos morais. Como outras formas de filosofia experimental, ela envolve a coleta de dados por métodos experimentais, e o uso destes dados para fundamentar, minar ou revisar teorias filosóficas. No presente caso, as teorias dizem respeito à natureza do raciocínio e do juízo moral; à extensão e às fontes das obrigações morais; à natureza de uma pessoa e de uma vida boa; e mesmo ao escopo e

---

* ALFANO, M.; LOEB, D.; PLAKIAS, A. Experimental Moral Philosophy. *In*: ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy.** Winter Edition. Stanford, CA: The Metaphysics Research Lab. Disponível em: https://plato.stanford.edu/archives/win2018/entries/experimental-moral/. Acesso em: 24 jul. 2022.

The following is the translation of the entry on Experimental Moral Philosophy by Mark Alfano, Don Loeb and Alexandra Plakias, 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/experimental-moral/. 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 a https://plato.stanford.edu/entries/experimental-moral/. 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.

natureza da própria teoria moral. Este verbete começa com uma breve consideração da história dos usos de dados empíricos na teoria moral, e então considera a questão acerca de qual a característica distintiva da filosofia moral **experimental**, se é que existe uma tal característica. Como deveríamos distinguir a filosofia moral experimental de trabalhos relacionados em psicologia moral empírica? Após discutir algumas da estratégias de resposta a esta questão, o verbete examina dois dos principais projetos em filosofia moral experimental, e então discute algumas das áreas mais proeminentes de pesquisa neste campo. Como veremos, em alguns casos a filosofia moral experimental abriu novos caminhos de investigação, enquanto em outros ela influenciou debates de longa data em teoria moral.

1. Introdução e História

A ideia de que nossos juízos morais são uma importante fonte de informação sobre as origens e a justificação das normas morais remonta à Grécia antiga, ou talvez até a um passado ainda mais remoto. Heródoto reconta a história de um imperador persa, Dario, que convidou membros gregos de sua corte "e perguntou que preço os convenceria a comer os cadáveres de seus pais. Eles responderam que não havia um preço pelo qual fariam isto". Foi então que Dario convocou membros de um grupo diferente "e perguntou… o que os faria queimar os pais depois de mortos. Os indianos gritaram que ele não deveria sequer mencionar um ato tão terrível". Heródoto conclui que histórias como esta provam, como o poeta Píndaro escreveu, que "o costume é o rei de todos". E assim ele fornece um exemplo do argumento do desacordo moral a favor do relativismo. Em *Outlines of Skepticism*, Sexto Empírico também enfatiza que descobertas empíricas podem desestabilizar nossa confiança no acordo moral universal.

> Mesmo que em alguns casos não possamos ver uma anomalia, deveríamos dizer que é possível que também haja disputa sobre eles… De modo similar, se não tivéssemos conhecido, por exemplo, o costume dos egípcios de se casar com suas irmãs, teríamos afirmado erroneamente que todos concordam que você não deve se casar com sua irmã.

Embora o uso de observação empírica em teoria moral tenha uma longa história, o movimento contemporâneo conhecido como filosofia experimental remonta a apenas algumas décadas. O movimento atual deve seu início aos trabalhos de Stephen Stich e Jonathan Weinberg (2001), e Joshua Knobe (2003). Mas talvez a primeira instância de filosofia experimental seja o *Truth as Conceived by Those Who Are Not Professional Philosophers* (NAESS, 1938). Embora não tenha havido uma aceitação uniforme do movimento por parte dos filósofos contemporâneos, veremos que há razões para pensar que a evidência experimental pode ter um papel, significância e importância distintiva na filosofia e teorização moral.

A relação entre a filosofia mais tradicional e o trabalho experimental é instrutiva, e traz à tona algumas tensões presentes na filosofia e teoria moral: nomeadamente, a moralidade é ao mesmo tempo prática e normativa, e estes dois aspectos informam e limitam a medida em que ela responde à psicologia humana.

Na medida em que a moralidade é prática, ela deve ser acessível e viável a agentes como nós: se uma teoria é muito exigente, ou depende de intuições, juízos, motivações ou capacidades que as pessoas não possuem (ou, o que é ainda pior, nem podem possuir), então podemos descartá-la com base nesse mesmo fato. Por outro lado, a moralidade também é normativa: seu objetivo não é apenas descrever o que de fato pensamos ou fazemos, mas guiar nossa prática. Por esta razão, aos resultados experimentais que alegam mostrar que alcançar e expressar virtudes de forma confiável em uma ampla variedade de situações é algo difícil (veja a seção 3.1, abaixo, para uma discussão desta literatura), alguns filósofos respondem apontando que uma teoria não é falsificada pelo fato de que as pessoas nem sempre julgam corretamente, ou agem pelas razões corretas. Tudo que este fato mostra é que as pessoas muitas vezes agem de forma moralmente deficiente. Retornaremos a estes problemas ao final deste verbete, quando discutirmos críticas à filosofia moral experimental. Por enquanto, eles são mencionados para ilustrar que é uma questão em si mesma controversa até que ponto a filosofia moral experimental ameaça as abordagens filosóficas tradicionais. Alguns filósofos morais veem a si mesmos como derivando princípios morais *a priori*, sem apelar a fatos contingentes sobre a psicologia humana. Outros veem a si mesmos como trabalhando dentro de uma tradição que remonta pelo menos até Aristóteles, e que concebe a ética como o estudo do florescimento humano. Estes filósofos não necessariamente abraçaram a filosofia moral experimental. Mas muitos veem seus projetos como resultados das teorias morais naturalistas desenvolvidas por Aristóteles, Hume e outros.

1.1. O Que São Experimentos?

Como os exemplos discutidos acima revelam, vários tipos de evidência empírica são úteis para a teorização moral (veja, também, os verbetes sobre **psicologia moral empírica**, **justiça distributiva empírica**, e **psicologia empírica e caráter**). Observações e dados antropológicos há muito desempenham um papel na filosofia moral. Os filósofos morais John Ladd e Richard Brandt, do século XX, investigaram o relativismo moral parcialmente através da condução de suas próprias etnografias em comunidades nativas americanas. Brandt escreve: "Temos… uma questão que afeta a verdade do relativismo ético, e que a antropologia concebivelmente pode nos ajudar a resolver. Será que a evidência etnológica apoia a perspectiva de que "pessoas qualificadas" podem discordar quanto à atitude ética?". Mas ele observa que "alguns tipos de material antropológico não nos ajudarão – em particular as informações isoladas sobre diferenças interculturais com respeito a opinião ou atitude ética" (1952, p. 238). Eis uma advertência frequentemente citada por filósofos envolvidos com pesquisa empírica: é importante que filósofos participem do design do experimento e da coleta de dados empíricos, porque existem certas questões que têm de ser abordadas para que os dados tenham aplicações filosóficas. No presente caso, uma destas questões é se desacordos morais envolvem diferenças com respeito a crenças factuais, ou outras diferenças não morais. Barry Hallen (1986, 2000) conduziu uma série de entrevistas e etnografias entre os iorubás, investigando conceitos avaliativos centrais e linguagem relacionada à epistemologia, estética e valores morais. Hallen foi motivado por questões sobre a indeterminação da tradução, mas seu trabalho é um exemplo de como entrevistas minuciosas podem informar investigações de conceitos filosóficos. Embora estes exemplos mostrem que a etnografia tem um papel valioso a desempenhar na teoria filosófica, o restante deste verbete se concentrará primariamente em experimentos. Experimentos paradigmáticos envolvem a atribuição randomizada de objetos ou pessoas de uma amostra representativa a variadas condições, seguida pela comparação estatística dos resultados para cada condição. A variação nas condições é por vezes chamada de "manipulação". Por exemplo, um experimentador pode tentar selecionar uma amostra representativa de pessoas, distribuí-las randomicamente entre o grupo que encontra e o grupo que não encontra uma moeda de dez centavos na caixa de devolução de moedas de um telefone público, e então medir o grau em que eles ajudam alguém que consideram estar em necessidade (ISEN; LEVIN, 1972). Encontrar ou não encontrar a moeda é a condição; o grau de ajuda é a variável de resultado.

Embora verdadeiros experimentos sigam este procedimento, outros tipos de estudos permitem a atribuição não randomizada a condições. Astrônomos, por exemplo, por vezes falam de experimentos naturais. Nós não estamos em posição de afetar os movimentos dos corpos celestes, mas podemos observá-los sob uma variedade de condições, por longos períodos e com uma variedade de instrumentos. Do mesmo modo, embora seja improvável que a investigação antropológica de crenças e práticas morais de diferentes culturas envolva a manipulação de variáveis nas vidas dos membros da cultura, tais estudos são valiosos e empiricamente respeitáveis. A despeito disto, estes estudos podem no máximo demonstrar correlação, e não causação, de modo que seu valor evidencial é menor do que o de verdadeiros experimentos. A maioria das pesquisas publicadas em filosofia experimental envolve verdadeiros experimentos.

Mesmo dentro da categoria dos experimentos, encontramos muita diversidade com respeito a inputs, métodos de medição e outputs. Os experimentos que acabamos de descrever usam medição e manipulação do comportamento: encontrar a moeda de dez centavos é o input, ajudar é o resultado medido. Outros experimentos não medem comportamentos, e sim juízos e intuições. Isto pode ser feito por meio de um inquérito ou por outra forma de autorrelato ou relato de respondentes, nos quais sujeitos respondem explicitamente a alguma questão, situação ou dilema. Estudos que medem juízos podem usar manipulações da condição em que o sujeito ajuíza, ou podem procurar por correlações entre juízos e algum outro fator, como atividade cerebral, resposta emocional, tempo de reação, atenção visual e assim por diante (STROHMINGER *et al.,* 2014).

Os métodos experimentais disponíveis também mudaram ao longo do tempo. Os inquéritos têm sido o método dominante da filosofia experimental nas últimas décadas, mas a tecnologia pode mudar isto: avanços na realidade virtual ou aumentada significam que filósofos agora podem imergir pessoas em dilemas morais como o exemplo do violinista, um experimento de pensamento proposto por Thompson (1971), e as diferentes versões do problema do bonde. Filósofos interessados nos correlatos neurais de juízos morais podem usar estimulação magnética transcraniana (EMT) para investigar os efeitos de aumentar ou diminuir a atividade em certas áreas do cérebro. Até mesmo os métodos de inquérito avançaram graças à tecnologia. A onipresença dos smartphones permite que pesquisadores tenham acesso a pessoas em tempo real, e solicitem relatos acerca de seu humor (*vide* seção 3.2, abaixo, para uma discussão de inquéritos relacionados à felicidade e ao humor).

Se a filosofia moral experimental tem de usar verdadeiros experimentos, ou se ela pode incluir estudos ou mesmo etnografias e outras formas de dados qualitativos, é uma questão parcialmente terminológica sobre como definir o campo e sobre se o distinguimos da psicologia moral empírica, que é um programa de pesquisa intimamente relacionado. Como veremos abaixo, contudo, uma diversidade de métodos e temas é um auxílio importante para a filosofia moral experimental responder aos seus críticos.

1.2 Quais Tipos de Questões e Dados?

Assim como a filosofia experimental (embora o interesse desta seja mais geral), a filosofia moral experimental está interessada em nossas intuições e juízos sobre experimentos filosóficos de pensamento e dilemas morais. Mas a pesquisa nesta área também diz respeito à arquitetura cognitiva subjacente a nossos juízos morais, ao processo desenvolvimental que dá origem a juízos morais, à neurociência dos juízos morais, e a outros campos relacionados.

Experimentos **diretos** investigam se uma afirmação sustentada (ou negada) por filósofos é corroborada ou falsificada. Isto pode significar investigar uma intuição ou investigar se ela é tão amplamente compartilhada quanto os filósofos dizem, ou pode significar investigar a afirmação de que certo comportamento ou traço é generalizado ou que dois fatores covariam. Por exemplo, encontramos filósofos que afirmam que é errado prender uma pessoa inocente para evitar um motim, que uma vida boa tem de ser autêntica, e que juízos morais são intrinsecamente motivadores. A pesquisa experimental pode ser (e tem sido, como veremos abaixo) aplicada a cada uma destas afirmações.

Experimentos **indiretos** consideram a natureza de alguma capacidade ou juízo: por exemplo, se certos tipos de dilemas morais envolvem áreas particulares do cérebro; como crianças desenvolvem a capacidade de empatia; e se a distinção moral/convencional é universal entre as culturas. Estas afirmações têm relevância filosófica na medida em que nos ajudam a compreender a natureza do juízo moral.

Além das distinções envolvendo os tipos de dados e quão diretamente eles são relevantes para as questões em jogo, também podemos distinguir entre ética aplicada experimental, ética normativa experimental e metaética experimental. A primeira envolve as variáveis que influenciam juízos sobre problemas práticos particulares, por exemplo como um carro autônomo deveria responder a dilemas

de sacrifício (BONNEFON; SHARRIFF; RAHWAN, 2016). A segunda envolve investigações acerca de como devemos nos comportar, agir e julgar, e também nossas intuições sobre responsabilidade moral, caráter e o que constitui uma vida boa. A terceira envolve debates sobre realismo e antirrealismo moral. Neste verbete, nos concentramos nas duas últimas.

Em muitos destes casos, é difícil traçar a linha entre filosofia moral experimental e suas vizinhas: o que distingue a filosofia moral experimental da psicologia moral empírica? Qual a diferença entre a filosofia moral experimental e a psicologia ou a neurociência que investiga a moralidade? Qual a diferença entre a filosofia moral experimental e a metaética? Podemos tentar responder a estas questões apontando se os experimentadores são treinados como filósofos ou como cientistas, mas muito do trabalho nestas áreas é feito através de times que envolvem a colaboração tanto de filósofos como de cientistas sociais ou neurocientistas. Além disto, alguns filósofos trabalham em departamentos de direito e psicologia, e programas de pós-graduação oferecem cada vez mais treinamento interdisciplinar. Outra estratégia seria olhar para a literatura com a qual o trabalho relevante se engaja; muitos psicólogos (por exemplo, Haidt e Greene) que investigam o juízo moral situam seus argumentos no debate entre kantianos e humeanos, de modo que o engajamento com a tradição filosófica pode ser usada como um critério para distinguir a filosofia moral experimental da psicologia moral experimental. Isto posto, a incapacidade de traçar uma distinção precisa entre a filosofia moral experimental e áreas adjacentes de investigação pode não ser particularmente importante. Filósofos experimentais muitas vezes ressaltam o fato de que a divisão disciplinar entre a filosofia e a psicologia é um fenômeno relativamente recente. Escritores do início do século XX, como William James, se situaram em ambas as disciplinas, fazendo contribuições vitais para cada uma delas. Se existe um problema aqui, ele não é exclusivo à filosofia moral experimental. Por exemplo, será o trabalho em semântica que usa tanto linguística como filosofia analítica melhor entendido como linguística ou como filosofia da linguagem? Estes debates surgem, e podem ser largamente irrelevantes, para muitos programas de pesquisa que cruzam fronteiras disciplinares.

O restante deste verbete prossegue do seguinte modo. A seção 2 examina a pesquisa experimental sobre juízos e intuições morais. A seção descreve vários usos programáticos que têm sido atribuídos aos resultados experimentais, e então se volta para exemplos de pesquisa experimental sobre juízo e intuição moral, incluindo intuições sobre intencionalidade e responsabilidade, bem como a chamada "analogia linguística". A seção 3 discute resultados experimentais sobre tópicos

densos (ou seja, simultaneamente descritivos e normativos), incluindo caráter e virtude, bem-estar, e emoção. A seção 4 discute questões sobre desacordo moral e linguagem moral, sendo ambas importantes fontes de evidência no debate de longa data sobre objetividade moral. A seção 5 considera algumas objeções à filosofia moral experimental.

2. Intuições Morais e Análise Conceitual

Como indicado acima, um papel para experimentos em filosofia moral consiste em investigar nossas intuições morais sobre casos, e usar os resultados colhidos de tais investigações para guiar ou restringir nossa metafísica moral, semântica ou epistemologia. Filósofos frequentemente se apoiam em juízos da forma "nós diríamos que x" ou "intuitivamente, x" como dado ou evidência para uma teoria (*vide* Cappelen, 2012, e Deutsch, 2016, para uma discussão crítica deste ponto). Afirmações sobre o que nós diríamos ou sobre a resposta intuitiva a um caso são empiricamente testáveis, e um projeto (talvez dominante) em filosofia moral experimental tem sido o de testar tais afirmações por meio de uma investigação de nossas intuições e de juízos morais relacionados. Ao proceder deste modo, filósofos morais experimentais podem alcançar uma destas duas coisas: primeiro, eles podem testar as afirmações dos filósofos tradicionais sobre o que é e o que não é intuitivo; segundo, eles podem investigar as fontes de nossas intuições e juízos. Estas tarefas podem ser realizadas como parte de um programa positivo, que usa nossas intuições e juízos como inputs, e constrói teorias que as acomodam e explicam. Alternativamente, estas tarefas podem figurar em um programa negativo, que usa pesquisa experimental para minar o apelo tradicional à intuição como evidência, seja na filosofia moral ou na análise conceitual de forma mais ampla. O programa negativo pode proceder direta ou indiretamente. No primeiro caso, o programa envolve testar e refutar afirmações sobre as próprias intuições. No segundo, envolve desacreditar as fontes destas intuições, por meio da descoberta de que elas são influenciadas por fatores amplamente considerados não evidenciais ou não confiáveis. Abaixo, discutimos tanto o programa negativo como o programa positivo.

2.1. O Programa Negativo na Filosofia Moral Experimental

Os primeiros trabalhos em filosofia experimental sugeriram diferenças interculturais com respeito a intuições semânticas, epistêmicas e morais. Por exemplo, Machery, Mallon e Stich (2004) argumentaram que sujeitos do leste asiático tinham maior propensão em sustentar intuições descritivistas do que seus colegas ocidentais, que tendiam a aceitar teorias causais da referência. Haidt (2006) argumentou que a medida em que pessoas julgaram erradas a violações não danosas, como comer um animal de estimação falecido ou se envolver em incesto consensual entre irmãos, dependeu do status econômico das pessoas em questão. Estes resultados têm sido combinados para argumentar contra o valor evidencial de intuições ou contra a existência de fatos morais de todo em todo. Isto porque, presumivelmente, o erro moral não depende do status socioeconômico, do gênero ou da cultura da pessoa que faz o juízo moral.

A filosofia moral experimental negativa gera resultados que são posteriormente usados para desconsiderar o valor evidencial dos apelos à intuição. Por exemplo, Walter Sinnott-Armstrong (2008d), Eric Schwitzgebel e Fiery Cushman (2012), e Peter Singer (2005) seguiram recentemente esta linha de pensamento, argumentando que intuições morais estão sujeitas a influências situacionais normativamente irrelevantes (p. ex. efeitos de ordem). Já Feltz e Cokely (2009) e Knobe (2011) documentaram correlações entre intuições morais e diferenças individuais que (presumivelmente) são normativamente irrelevantes (por exemplo, extroversão). Estes resultados podem justificar o ceticismo sobre as intuições morais, ou pelo menos sobre algumas classes de intuições ou intuidores.[67]

Os estudos que acabamos de mencionar envolvem resultados que sugerem que o conteúdo das intuições varia ao longo de alguma dimensão normativamente irrelevante. Outra fonte de evidência para o programa negativo advém de resultados acerca dos mecanismos cognitivos subjacentes, que geram as intuições. Por exemplo, estudos que sugerem que intuições deontológicas são motivadas por aversões emocionais à prejudicar a outros têm sido usados para argumentar que

---

[67] Kwame Anthony Appiah (2008) e Peter Singer (2005) são outros que proeminentemente expressam interesse pela relevância de resultados experimentais como estes para a confiança dos filósofos nas intuições morais.

devemos desconsiderar nossas intuições deontológicas em favor de princípios consequencialistas. Esta ideia é discutida em maior detalhe abaixo (*vide* SINGER, 2005; WEIGMAN, 2017).

É difícil traçar a distinção entre filosofia moral experimental negativa e positiva. Em parte, isto ocorre porque o programa negativo muitas vezes desconsidera classes ou tipos particulares de intuições em favor de outras que, supostamente, são mais confiáveis. Singer, por exemplo, oferece um argumento antideontológico como parte de seu programa negativo, na medida em que seu argumento recorre às origens emocionais das intuições deontológicas para desconsiderá-las. Entretanto, sua posição também se enquadra no programa positivo, dado que ele posteriormente argumenta a favor da superioridade de intuições consequencialistas. Deste modo, a diferença entre os dois programas não reside nos dados ou nos tipos de questões investigadas, mas em como os dados são usados. A diferença depende de se os dados são vistos como desacreditando apelos filosóficos tradicionais à intuição, ou se são vistos como uma adição aos apelos filosóficos tradicionais à intuição, por ajudar filósofos a distinguir entre intuições confiáveis, não confiáveis e suas fontes. Na próxima seção, consideramos usos positivos de intuições como evidência.

2.2. O Programa Positivo na Filosofia Moral Experimental

Outros filósofos são mais otimistas sobre o resultado de investigações experimentais acerca do juízo e da intuição moral. Por exemplo, Knobe usa investigações experimentais dos fatores determinantes dos juízos morais para identificar os contornos de conceitos filosoficamente ressonantes e os mecanismos ou processos subjacentes ao juízo moral. Ele argumentou a favor da influência generalizada de considerações morais nos conceitos da psicologia de senso comum (*vide* KNOBE, 2009, 2010; PETTIT; KNOBE, 2009). Para isto, ele alegou, dentre outras coisas, que o conceito de uma ação intencional é sensível à valência avaliativa previsível das consequências da ação relevante (KNOBE, 2003, 2004b, 2006).[68] Em um artigo recente, Knobe (2006) mostra que somente dez por cento da filosofia experimental tem como objetivo apoiar ou minar alguma análise conceitual já

[68] Esta e outras pesquisas relacionadas serão discutidas em maior detalhe na subseção 2.3.

proposta. Pelo contrário, a maior parte dos artigos já publicados tem como objetivo esclarecer os processos cognitivos e afetivos subjacentes ao fenômeno relevante.

Outra linha de pesquisa remonta a Joshua Greene e seus colaboradores (GREENE *et al*., 2001, 2004, 2008; GREENE, 2008), que investigaram as bases neurais dos juízos morais consequencialistas e deontológicos. Greene e seus colaboradores suscitaram as intuições de sujeitos acerca de uma variedade de problemas do bonde enquanto estes estavam dentro de uma scanner RMf. Os problemas do bonde são casos que apresentam um dilema em que um bonde está acelerando em direção a cinco indivíduos, que serão todos mortos a menos que o bonde seja desviado para a outra direção, contendo apenas um indivíduo. Os pesquisadores descobriram que, quando apresentados a casos em que o bonde é desviado ao apertar um botão, a maioria dos sujeitos concordou que a ação correta era desviar o bonde dos cinco, em direção ao indivíduo isolado. Quando, ao invés disto, o caso envolvia empurrar uma pessoa de uma ponte para que caísse em frente ao bonde de forma a bloqueá-lo, os sujeitos estavam menos propensos a julgar como moralmente aceitável o sacrifício de uma pessoa. Além disto, Greene descobriu que sujeitos que julgaram que empurrar a pessoa era permissível demoraram mais para formar seu juízo, o que sugere que eles tiveram de superar intuições conflitantes para fazê-lo. Esta descoberta é reforçada pelo fato de que, quando os sujeitos consideraram empurrar a pessoa da ponte, seus cérebros revelaram aumento de atividade em áreas associadas com a agregação e modulação de sinais de valor (p. ex. córtex pré-frontal ventromedial e córtex pré-frontal dorsolateral). Greene conclui que nossa aversão a empurrar a pessoa se deve a uma resposta emocional ao pensamento de causar dano físico a alguém, enquanto nossa disposição de apertar o botão se deve ao cálculo racional de que salvar cinco pessoas é preferível a deixar cinco pessoas morrerem para poupar uma vida. Greene e Singer usam estas descobertas como base para desacreditar intuições deontológicas e defender o consequencialismo, pois este repousa sobre intuições que advém de uma fonte que consideramos ser geralmente confiável e apropriadamente usada como a base para raciocínio moral. É importante notar, contudo, que este argumento pressupõe que (pelo menos nestes casos) reações emocionais são necessariamente irracionais ou arracionais. Filósofos que não são simpáticos a esta perspectiva das emoções não precisam seguir Greene e Singer até suas últimas conclusões (RAILTON, 2014).

Uma abordagem relacionada tem como objetivo a identificação das características às quais intuições sobre conceitos filosoficamente importantes são

sensíveis. Sripada (2011) pensa que o papel adequado às investigações experimentais das intuições morais não é o de identificar os mecanismos subjacentes a estas intuições. Ele afirmou que tal conhecimento tem pouca relevância para a teorização filosófica. Antes, o relevante é investigar, considerando caso a caso, as características às quais as pessoas estão respondendo quando elas têm as intuições relevantes. De acordo com esta perspectiva, pessoas (incluindo filósofos) podem prontamente identificar se elas têm uma dada intuição, mas não porque a tem. Eis um exemplo extraído do debate sobre determinismo e livre-arbítrio: pensa-se que casos envolvendo manipulação minam intuições compatibilistas – intuições que apoiam a ideia de que o determinismo é compatível com "o tipo de liberdade necessária para a responsabilidade moral" (PEREBOOM, 2001). Tais casos envolvem a descrição de uma vítima involuntária sendo sub-repticiamente manipulada para ter e refletidamente defender uma motivação para realizar alguma ação. Críticos do compatibilismo dizem que tais casos satisfazem critérios compatibilistas para a responsabilidade moral, mas que, intuitivamente, os agentes em questão não são moralmente responsáveis (PEREBOOM, 2001). Entretanto, através de análise de mediação e modelagem de equações estruturais, Sripada (2011) faz uma defesa plausível de que as pessoas sentem em alguma medida que o indivíduo manipulado não é moralmente responsável porque julgam que ele na verdade não satisfaz os critérios compatibilistas.[69] Deste modo, ao determinar a quais aspectos do caso as intuições filosóficas estão respondendo, talvez seja possível resolver questões que seriam de outro modo intratáveis.

2.3. Um Exemplo: Intencionalidade e Responsabilidade

Desde o seminal artigo de Knobe (2003), filósofos experimentais têm investigado os padrões complexos nas disposições das pessoas de fazer juízos sobre noções morais (louvável, culpável, responsabilidade), atitudes cognitivas (crença, conhecimento, recordação), atitudes motivacionais (desejo, favorecimento,

---

[69] A análise de mediação tenta determinar se uma variável (a preditora) afeta uma segunda variável (o resultado) ao influenciar uma terceira variável, a mediadora (BARON; KENNY, 1986). A modelagem de equações estruturais permite que o analista avalie e compare vários modelos da relação entre preditores, resultados, mediadores e moderadores (KLINE, 2005).

aceitação) e traços de caráter (compaixão, insensibilidade) no contexto de violações de, ou conformidade a, várias normas (morais, prudenciais, estéticas, legais, convencionais, descritivas).[70] No exemplo original de Knobe, os participantes primeiro leem uma descrição de um cenário envolvendo uma escolha: o protagonista é apresentado a uma potencial medida (que objetiva aumentar os lucros) que resultaria em um efeito colateral (prejudicar ou ajudar o meio ambiente). Em seguida, o protagonista rejeita explicitamente a consideração do efeito colateral, e escolhe seguir em frente com a medida. A medida tem o resultado anunciado: tanto o efeito primário como o efeito colateral ocorrem. Pede-se aos participantes da pesquisa que atribuam intencionalidade ou uma atitude (ou um traço de caráter, no caso de um trabalho posterior de Robinson *et al.*, 2013) ao protagonista. Knobe descobriu que os participantes estavam significativamente mais inclinados a indicar que o protagonista provocou intencionalmente o efeito colateral quando este efeito era percebido como ruim (prejudicar o meio ambiente) do que quando ele era percebido como bom (ajudar o meio ambiente). Este efeito foi replicado dezenas de vezes, e seu escopo tem sido consideravelmente expandido de atribuições de intencionalidade após violações de uma norma moral para atribuições de propriedades variadas após violações de uma ampla variedade de normas.

O objetivo primário dos intérpretes deste corpo de evidência é criar um modelo capaz de predizer quando a assimetria na atribuição surgirá. Os objetivos secundários são explicar, da forma mais sistemática possível, a razão destes efeitos ocorrerem, e determinar em que medida a assimetria na atribuição pode ser considerada racional. A figura 1, apresentada pela primeira vez aqui, modela como a resposta dos participantes a este tipo de vinheta são produzidas.

---

[70] Veja Nadelhoffer (2004, 2006), Knobe e Mendlow (2004), Knobe (2004a, 2004b, 2007), Pettit e Knobe (2009), Tannenbaum, Ditto e Pizarro (2007), Beebe e Buckwalter (2010), Beebe e Jensen (2012), Alfano, Beebe e Robinson (2012), Robinson, Stey e Alfano (2013).

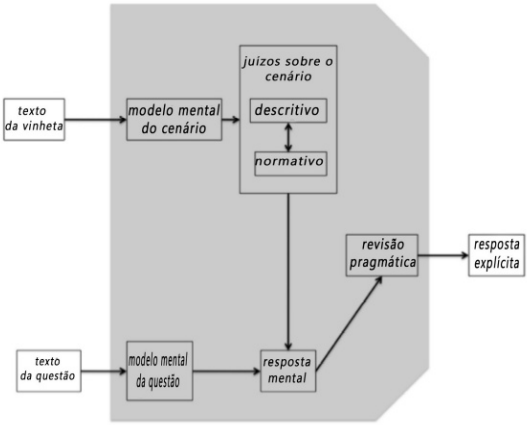


Figura 1: Modelo de resposta do participante a vinhetas de filosofia experimental

Neste modelo, as caixas representam construtos, as setas representam influências causais ou funcionais, e a área em cinza representa a mente do participante. Embora a mente do participante não seja diretamente observável, é o alvo da investigação. Em linhas gerais, a ideia é que primeiro um participante lê o texto da vinheta e forma um modelo mental do que acontece na história. Com base neste modelo (e quase certamente enquanto a vinheta ainda está sendo lida), o participante começa a interpretar; ou seja, ele começa a fazer juízos descritivos e normativos sobre o cenário, e em particular sobre os estados mentais e sobre os traços de caráter das pessoas envolvidas no mesmo. O participante então lê a questão proposta pelo experimentador, forma um modelo mental do que está sendo perguntado e, com base em seus juízos sobre o cenário, forma uma resposta para a questão. A resposta pode então ser pragmaticamente revisada (para evitar implicações indesejadas, para colocá-la em maior acordo com o que o participante pensa que o experimentador quer ouvir, etc.) e por fim registrada como uma resposta explícita a uma afirmação sobre as atitudes do protagonista (por exemplo, "ele provocou o efeito colateral intencionalmente", classificada em uma escala Likert)[71].

[71] Estas escalas são nomeadas em homenagem ao seu inventor, Rensis Likert (1932). Uma afirmação é apresentada ao participante, que, então, é solicitado a concordar ou discordar dela em uma escala numérica. Usualmente, as escalas vão de 1 a 7, 1 a 5, -3 a 3, ou -2 a 2. Os pontos finais da escala são quase sempre rotulados de "discordo fortemente" e "concordo fortemente". O ponto intermediário é muitas vezes rotulado como "nem concordo nem discordo". Por vezes, outros pontos da escala também são rotulados.

Um resultado que tem sido repetidamente replicado é que, quando a vinheta descreve uma violação de norma, sujeitos indicam concordar mais fortemente que a ação violadora da norma foi realizada intencionalmente. Esta descoberta poderia ser usada pelos defensores do programa negativo para atacar a coerência conceitual de nossa noção de responsabilidade moral. Mas filósofos morais experimentais que trabalham no programa positivo têm assumido a tarefa de explicar a assimetria, através da postulação de modelos de entidades inobserváveis que mediam, explicam e talvez até racionalizem a assimetria. No que segue, apresento uma discussão de várias tentativas deste tipo. Embora cada uma ofereça uma explicação diferente da assimetria, todas representam defesas, ou pelo menos reabilitações, de nossas intuições sobre intenção.

O Modelo da Competência Conceitual

Talvez o mais conhecido seja o modelo da competência conceitual, de Knobe, de acordo com o qual a assimetria surge no estágio do juízo. Nesta perspectiva, juízos normativos sobre a ação influenciam juízos descritivos sobre se ela foi intencional (ou desejada, esperada, etc.). Além disto, esta influência é considerada parte da própria concepção de intencionalidade (desejo, crença, etc.). Assim, de acordo com o modelo da competência conceitual, a assimetria nas atribuições é uma expressão racional da concepção comum de intencionalidade (desejo, crença, etc.), que tem um componente normativo.[72]

O Modelo do Viés Motivacional

O modelo do viés motivacional (ALICKE, 2008; NADELHOFFER, 2004, 2006) concorda que a assimetria tem origem no estágio do juízo, e que juízos normativos influenciam juízos descritivos. Entretanto, diferentemente do modelo da competência conceitual, considera-se que isto seja um viés, e não uma expressão de competência conceitual. Assim, neste modelo a assimetria nas atribuições é

---

[72] A ideia de que conceitos aparentemente preditivos e explicativos também podem ter um componente normativo não se originou inteiramente em Knobe. Bernard Williams ressaltou que virtudes e vícios têm este caráter misto.

uma distorção da concepção correta de intencionalidade (desejo, crença, etc.).

O Modelo do Eu Profundo

O **modelo da concordância com o eu profundo** (SRIPADA, 2010, 2012; SRIPADA; KONRATH, 2011) também localiza a origem da assimetria no estágio do juízo. Mas ele não reconhece uma influência (lícita ou ilícita) de juízos normativos sobre juízos descritivos. Ao contrário, defensores deste modelo afirmam que, ao acessar a ação intencional de um indivíduo, as pessoas não se atentam apenas à sua representação de um eu "superficial" do indivíduo: as expectativas do indivíduo, suas crenças sobre meios e fins, intenções de momento a momento e desejos condicionais. Elas também se atentam à sua representação do eu "profundo" do indivíduo, que abriga seus sentimentos, valores e princípios fundamentais. De acordo com este modelo, ao avaliar se alguém produziu intencionalmente um estado de coisas, pessoas determinam (tipicamente de forma inconsciente) se existe concordância suficiente entre sua representação do resultado que o agente produz e o que eles consideram ser o eu profundo do agente. Por exemplo, quando o presidente diz que não se importa minimamente em prejudicar ou ajudar o meio ambiente, pessoas atribuem a ele uma postura profundamente antiambiental. Quando o resultado é que ele prejudica o meio ambiente, está em acordo com seu eu profundo antiambiental. Quando, em contraste, o resultado é que ele ajuda o meio ambiente, está em desacordo com seu eu profundo antiambiental. Portanto, de acordo com o modelo da concordância com o eu profundo, a assimetria nas atribuições é uma expressão razoável da distinção psicológica de senso comum entre o eu profundo e o superficial.

O Modelo da Pragmática Conversacional

Diferentemente dos modelos discutidos até agora, o modelo da pragmática conversacional (ADAMS; STEADMAN, 2004, 2007) localiza a origem da assimetria no estágio da revisão pragmática. De acordo com este modelo, tanto nos casos envolvendo conformidade como naqueles envolvendo violação de normas os participantes julgam que o protagonista não agiu intencionalmente. Entretanto, quando chega o momento de contar ao experimentador o que eles pensam, os participantes não querem ser tomados como se estivessem sugerindo que o protagonista causador do dano é inocente. Assim, eles informam que o protagonista

agiu intencionalmente. O objetivo dos participantes é razoável. Portanto, de acordo com o modelo da revisão pragmática, a atribuição da assimetria é racional, embora seja fonte de equívocos.

O Modelo da Deliberação

De acordo com o modelo da deliberação (ALFANO; BEEBE; ROBINSON, 2012; ROBINSON; STEY; ALFANO, 2013; SCAIFE; WEBBER, 2013) a melhor explicação dos padrões complexos de evidência é que o primeiro estágio, o da formação de um modelo mental do cenário, difere entre vinhetas envolvendo violação da norma e conformidade com a norma. Quando o protagonista é informado de que a medida que ele comumente desejaria seguir viola uma norma, ele adquire uma razão para deliberar mais sobre o que fazer. Em contraste, ele não adquire tal razão quando é informado de que a medida se conforma a alguma norma. Os participantes tendem a pensar o protagonista como alguém que está deliberando sobre o que fazer quando, e somente quando, uma norma seria violada. Uma vez que a deliberação leva à formação de outros estados mentais, tais como crenças, desejos e intenções, esta diferença básica entre os modelos dos participantes acerca do que acontece na história flui para o resto da interpretação deles, levando à atribuição de uma assimetria. De acordo com o modelo da deliberação, portanto, a atribuição da assimetria se origina antes do que outros filósofos experimentais supõem, e se deve a um processo racional.

2.4. Outro Exemplo: A Analogia Linguística

Outro programa positivo estuda a estrutura e a forma das intuições morais, além de seu conteúdo, com o objetivo de usar estas características como fonte de informação para uma teoria das estruturas cognitivas subjacentes ao juízo moral e sua etiologia.

Baseando-se na teoria da linguística gerativa de Chomsky (1965), Rawls (1971) sugeriu que a cognição moral pode ser convenientemente modelada em nossa faculdade da linguagem. O próprio Chomsky defendeu este paralelo.

> Não duvido que tenhamos um senso moral natural [...] Ou seja, assim como de algum modo pessoas

> podem construir um sistema extraordinariamente rico de conhecimento da linguagem com base em uma experiência limitada e corrompida, pessoas também desenvolvem sistemas implícitos de avaliação moral, que são mais ou menos uniformes de pessoa para pessoa. Existem diferenças, e as diferenças são interessantes. Entretanto, em uma medida bem substancial, tendemos a fazer juízos comparativos e, ao que parece, fazemos isto de maneiras bem intrincadas e delicadas, que envolvem novos casos e acordos frequentes sobre novos casos… e fazemos isto com base em um contexto ambiental muito limitado disponível a nós. A criança ou o adulto não possui muita informação que permita a pessoa madura a construir um sistema moral que de fato se aplicará a uma ampla gama de casos, e, no entanto, isto acontece… onde quer que vejamos um sistema muito rico e intrincado se desenvolvendo de modo mais ou menos uniforme com base em condições de estímulo restritas, temos de assumir que existe um componente inato muito poderoso, muito rico e altamente estruturado operando de modo a criar o sistema altamente específico com base nos dados limitados disponíveis a ele. (CHOMSKY *apud* MIKHAIL, 2005).

Aqui, Chomsky aponta quatro similaridades entre o desenvolvimento do conhecimento linguístico e o desenvolvimento do conhecimento moral.

> **L1:** É quase inevitável que uma criança criada em uma comunidade linguística particular acabe falando um idioleto da linguagem local, a despeito da falta de instrução explícita adequada, da falta de reações negativas abrangentes aos erros, e dos erros gramaticais daqueles que a criaram.
>
> **M1:** É quase inevitável que uma criança criada em uma comunidade moral particular acabe julgando de acordo com um idioleto do código moral local, a despeito da falta de instrução explícita adequada, da

falta de reações negativas suficientes aos erros morais, e dos erros morais daqueles que a criaram.

L2: Embora exista grande diversidade entre linguagens naturais, existem restrições sistemáticas sobre as linguagens naturais possíveis.

M2: Embora exista grande diversidade entre moralidades naturais, existem restrições sistemáticas sobre moralidades naturais possíveis.

L3: Usuários da linguagem obedecem a muitas regras esotéricas que, tipicamente, eles próprios não são capazes de articular ou explicar, e que alguns sequer reconheceriam.

M3: Agentes morais julgam de acordo com regras esotéricas (tal como a doutrina do duplo efeito) que, tipicamente, eles próprios não são capazes de articular ou explicar, e que alguns sequer reconheceriam.

L4: Baseando-se em um vocabulário limitado, um falante pode produzir e compreender uma potencial infinidade de expressões linguísticas.

M4: Baseando-se em um vocabulário moral limitado, um agente pode produzir e avaliar uma classe muito ampla (embora talvez não infinita) de planos de ação, que são aptos para o juízo moral.

No que segue, explicamos e avaliamos cada um destes pares de afirmações.

L1/M1 se referem ao **argumento da pobreza de estímulo**, de Chomsky: apesar de receberem pouca instrução linguística e gramatical explícita, crianças desenvolvem a linguagem rapidamente e em tenra idade, e rapidamente passam a manifestar competência na aplicação de regras gramaticais complexas. Do mesmo modo, argumentam Mikhail e outros defensores da analogia, em tenra idade crianças adquirem conhecimento moral baseado em uma instrução explicita relativamente pequena. Entretanto, críticos da analogia apontam vários pontos de diferença.

Primeiro, crianças de fato recebem uma boa dose de instrução e correção moral explícita, e, em contraste com o caso linguístico, estas instruções muitas vezes assumem a forma de afirmações explícitas de regras morais: "não bata", "compartilhe", e assim por diante. Segundo, existe um debate acerca da idade em que crianças de fato manifestam competência moral. Paul Bloom (2013) (*vide* BLAKE; MCAULIFFE; WARNEKEN, 2014) argumentou que bebês manifestam tendências morais ainda com 3-6 meses. Mas outros argumentaram que crianças não são totalmente competentes, no que diz respeito a juízos morais, até os 8-12 anos, idade em que já receberam uma boa dose de instrução moral explícita e implícita (os mais famosos a defender esta tese são Kohlberg (1969) e Piaget (1970); veja, também, Turiel (1983) e Nucci (2002)). Um ponto relacionado diz respeito à habilidade de receber instruções. No caso da linguagem, é preciso que uma criança tenha algum conhecimento ou entendimento linguístico até mesmo para receber instrução: uma criança com nenhuma compreensão da linguagem não entenderá as instruções dadas a ela. No caso da moral, contudo, não é necessário que uma criança compreenda as regras morais para que seja instruída por elas, e assim existe uma menor necessidade de postular algum conhecimento ou entendimento inato aqui. Nichols *et al.* (2016) conduziu uma série de experimentos usando aprendizado estatístico e modelagem bayesiana para mostrar como crianças podem aprender regras complexas com input escasso. Por fim, embora crianças possam, inicialmente, adquirir os valores morais presentes em seu meio, elas por vezes mudam seus valores, ou desenvolvem seus próprios valores mais tarde na vida, de modos espontâneos e com pouco esforço consciente. Isto está em claro contraste com a linguagem: como qualquer aprendiz de uma segunda língua reconhecerá, adquirir uma nova língua mais tarde na vida requer esforço e, mesmo quando alguém é bem-sucedido em alcançar a fluência, a primeira língua raramente é perdida. Para encerrar, como Prinz (2008) apontou, crianças são punidas por violações morais, e isto pode explicar por que elas são mais rápidas em aprender regras morais do que regras gramaticais.

L2/M2 se referem à existência dos chamados **universais linguísticos**: estruturas presentes em todas as linguagens conhecidas. Por razões que discutiremos mais adiante, é controverso que isto seja verdadeiro para o caso da moral. Prinz (2008) e Sripada (2004) argumentaram que não existem universais morais livres de exceções, a menos que se expresse ou descreva as normas em questão de modo a torná-las vazias. Sripada usa o exemplo de "assassinar é errado" como uma norma vazia. Esta norma pode parecer uma candidata plausível a universal

moral. Mas quando percebemos que “assassinar” se refere à ação incorreta de matar, podemos ver que a norma não é informativa. Mikhail pode responder alegando que o fato de que todas as culturas reconhecem um subconjunto dos assassinatos intencionais como moralmente errado é por si só significativo, mesmo se as culturas diferem no modo como circunscrevem este subconjunto, e que a analogia relevante aqui seria com a existência de categorias como “sujeito”, “verbo” e “objeto” em todas as linguagens, ou com características como a recursividade, que está presente na gramática de todas as linguagens naturais.

L3/M3 se referem aos padrões manifestados por intuições gramaticais e morais. No caso da linguagem, embora falantes nativos possam reconhecer construções gramaticais e agramaticais com relativa facilidade, eles tipicamente não conseguem articular as regras subjacentes a estas intuições. De acordo com a analogia em questão, ele é verdade com respeito ao caso da gramática moral: produzimos intuições morais sem esforço, mas somos incapazes de explicar as regras subjacentes a elas. De fato, as regras subjacentes aos juízos em ambos os casos podem ser bem esotéricas, e pode ser difícil para falantes nativos aprendê-las e entendê-las, como pode ser facilmente atestado por qualquer um que tenha ensinado a doutrina do duplo efeito para estudantes de nível introdutório.

L4/M4 se referem ao fato de que podemos combinar expressões linguísticas com outras expressões linguísticas para criar expressões e frases novas e cada vez mais complexas. Esta propriedade é conhecida como **recursividade**, e está presente em todas as linguagens naturais conhecidas (de fato, em um artigo de 2002, Chomsky sugere que a recursividade talvez seja o único universal linguístico; veja Hauser, Chomsky, e Fitch (2002)). Eis um exemplo de como expressões podem ser combinadas com outras expressões para formar expressões mais complexas:

> O gato malhado → o gato malhado (que o cachorro perseguiu) → o gato malhado (que o cachorro [que o conglomerado de criadouros queria] perseguiu) → o gato malhado (que o cachorro [que o conglomerado de criadouros {que estava falido} queria] perseguiu)

Do mesmo modo, juízos morais podem ser combinados com outros juízos morais para produzir novos juízos morais (HARMAN, 2008, p. 346). Por exemplo:

> É errado fazer *x* →É errado coagir alguém a fazer *x* → É errado persuadir alguém a coagir alguém a fazer *x*.

O ponto é duplo: primeiro, podemos criar descrições de ações complexas; segundo, podemos avaliar ações novas e complexas, e responder com intuições relativamente automáticas acerca da permissibilidade gramatical ou moral. Mikhail (2011, p. 43-48) usa evidência experimental envolvendo juízos sobre problemas do bonde para argumentar que nossos juízos morais são gerados pela imposição de uma estrutura deôntica à nossa representação dos aspectos causais e avaliativos da ação sob consideração. Ele aponta para uma variação do argumento da pobreza de estímulo que chama de "**pobreza do estímulo perceptivo**" (MIKHAIL, 2009, p. 37): quando somos confrontados com uma situação moral particular, extraímos inferências complexas sobre a ação e o agente com base em relativamente poucos dados. Mikhail (2010) usa este argumento contra modelos que veem juízos morais como intuições guiadas por emoções.

> Embora cada um destes juízos morais rápidos, intuitivos e em larga medida automáticos seja ocasionado por um estímulo identificável, a estrutura superficial do próprio estímulo não revela de qualquer modo óbvio como o cérebro interpreta estas descrições de ações complexas e como atribui um status deôntico a cada uma delas. Ao contrário, um passo intermediário deve ser postulado: algum tipo de avaliação intuitiva que é imposta ao estímulo antes de qualquer resposta deôntica a ele. Consequentemente, um modelo perceptivo simples, tal como aquele implícito no influente modelo do juízo moral proposto por Haidt (2001), parece inadequado para explicar estas intuições. Este ponto pode ser ilustrado pela atenção ao elo não analisado entre a situação evocada e a resposta intuitiva no modelo de Haidt.

Diferentemente do argumento tradicional da pobreza de estímulo, esta versão não se apoia em evidência desenvolvimental. Ela se apoia em nossa habilidade de avaliar cenários complicados, como as variações dos problemas do bonde, de forma rápida e sem esforço, e em seguida emitir juízos normativos. O

passo intermediário postulado é como o apelo inconsciente às regras da gramática, e a evidência para este passo tem de vir de experimentos mostrando que nós de fato temos as intuições relevantes, e que elas obedecem a padrões predizíveis.

A analogia linguística se apoia em evidência experimental acerca da natureza e padrão de nossas intuições morais, e indica um importante papel para experimentos na teoria moral. Se a analogia se sustenta, então uma teoria moral adequada tem de ser baseada nos juízos dos juízes morais competentes, assim como regras gramaticais são construídas com base nos juízos dos falantes competentes. A coleta sistemática de tais juízos será uma tarefa urgente. A analogia também sugere um tipo de racionalismo acerca do juízo moral, pois postula que nossos juízos resultam da aplicação inconsciente de regras a casos. Por fim, a analogia tem implicações metaéticas. Assim como só podemos avaliar a gramaticalidade de um proferimento relativamente a uma linguagem, talvez só possamos avaliar a moralidade de uma ação relativamente a um sistema moral específico. Como individuar moralidades e quantas existem (assumindo que existem várias), é uma questão para a investigação empírica posterior, e aqui, uma vez mais, a evidência experimental terá um papel crucial.

3. Caráter, Bem-estar e Emoção

Até 1950, a filosofia moral moderna concentrou-se principalmente no consequencialismo e no deontologismo. A revitalização da ética da virtude levou a um interesse renovado pelas virtudes e vícios (por exemplo, honestidade, generosidade, justiça, desonestidade, mesquinhez, injustiça), pela *eudaimonía* (muitas vezes, traduzida como "felicidade" ou "florescimento") e pelas emoções. Nas últimas décadas, recorreu-se a trabalhos experimentais em psicologia, sociologia e neurociência para dar fundamentação empírica a perspectivas filosóficas sobre estes temas.

3.1. Caráter e Virtude

Uma virtude é uma disposição complexa que compreende subdisposições de perceber, interpretar, pensar, desejar e agir de maneiras características. Por exemplo, ser generoso é (dentre outras coisas) estar disposto a perceber ocasiões

propícias à doação, a interpretar caridosamente a sinais sociais ambíguos, a desejar dar às pessoas coisas que elas querem, precisam ou apreciariam, a deliberar bem sobre o que elas querem, precisam ou apreciariam, e a agir com base nesta deliberação. Manifestações desta disposição são observáveis e, portanto, propícias para a investigação empírica. Eticistas da virtude das últimas décadas foram algumas vezes otimistas com respeito à distribuição da virtude na população. Alasdair MacIntyre, por exemplo, afirma que "sem alusão ao lugar que justiça e injustiça, coragem e covardia ocupam na vida humana, muito pouco será genuinamente explicável" (1984, p. 199). Julia Annas (2011, p. 8-10) afirma que "no momento em que refletimos sobre as virtudes, nós já temos alguma". Linda Zagzebski (2010) fornece uma semântica "exemplarista" para termos de virtude que funciona apenas se de fato existirem muitas pessoas virtuosas.

Partindo do *Varieties of Moral Personality* (1993), de Owen Flanagan, filósofos começaram a questionar se os resultados empíricos da psicologia social eram inconsistentes com a estrutura da ação humana pressuposta pela teoria da virtude. Esta abordagem concebe as pessoas como possuindo traços mais ou menos fixos de caráter, que sistematicamente comandam sua percepção, cognição, emoção, raciocínio, decisão e comportamento. Por exemplo, uma pessoa generosa está inclinada a perceber e buscar por oportunidades de fazer doações supererrogatórias aos outros. A pessoa generosa também está inclinada a pensar sobre o que seria e o que não seria apreciado pelos potenciais favorecidos, a sentir o ímpeto de doar e o brilho da satisfação após fazê-lo, a deliberar de forma efetiva sobre quando, onde e como doar a quem, a alcançar decisões firmes baseadas em tal deliberação, e a seguir com estas decisões uma vez que tenham sido tomadas. Outros traços devem se encaixar no mesmo padrão, estruturando a percepção, cognição, motivação e ação daqueles que os possuem. Resultados famosos em psicologia social, como o experimento do Bom Samaritano, de Darley e Batson (1973), parecem contradizer esta perspectiva da conduta moral humana. Quando uma pessoa ajuda outra em necessidade, pode ser que o faça simplesmente porque não está com pressa, ao invés de fazê-lo porque está expressando um traço fixo como generosidade ou compaixão.

Na abordagem da teoria da virtude, não é necessário assumir que as pessoas já são virtuosas. Entretanto, assume-se que as pessoas são pelo menos potencialmente sensíveis às considerações que uma pessoa virtuosa normalmente perceberia e levaria em conta. Flanagan (1993), seguido por Doris (1998, 2002), Harman (1999, 2000) e Alfano (2013) criou dificuldades para esta abordagem ao apontar evidências da psicologia social que sugerem que boa parte do pensamento,

sentimento e ação das pessoas é na verdade predito por (e consequentemente sensível a) fatores situacionais que de modo algum parecem contar como razões – nem mesmo más razões ou tentações para o vício. Estes fatores incluem a influência de estímulos do ambiente (sons, odores, iluminação, etc.), indutores de humor positivo ou negativo que são aparentemente triviais e normativamente irrelevantes, ordem de apresentação dos estímulos, e uma variedade de efeitos de enquadramento e pré-ativação, muitos dos quais são examinados em Alfano (2013: 40–50).[73] Vale a pena enfatizar o quão profundo é o problema colocado por estes estudos. O ponto não é que eles sugerem que a maioria das pessoas não é virtuosa (embora isto também seja sugerido). O ponto é que eles minam a inteira estrutura a partir da qual as pessoas são concebidas como cognitivamente e motivacionalmente sensíveis às ações. Alguém que falhou em agir virtuosamente porque cedeu à tentação pode ser compreendido a partir da abordagem da teoria da virtude. Alguém que falhou em agir virtuosamente porque acabou de ser subliminarmente pré-ativado com frio físico, que, por sua vez, é metaforicamente associado com frieza social, não tem qualquer lugar na abordagem da teoria da virtude. Estes tipos de efeitos nos levam a reformular completamente nossa noção de ação e pessoa (DORIS, 2009).

Estimativas iniciais sugeriram que variáveis de diferença individual tipicamente explicam menos que 10% da variação no comportamento das pessoas (MISCHEL, 1968), enquanto, como Funder e Ozer (1983) apontaram, fatores situacionais podem explicar menos que 16%.[74] Evidências agregadas mais recentes indicam

---

[73] Owen Flanagan (1991) considerou algumas das mesmas evidências antes de Doris e Harman, mas ele hesitou em tirar as mesmas conclusões pessimistas deles sobre a ética da virtude.

[74] Quando se trata de explicar a variação no comportamento, a ideia básica é que a análise estatística de resultados experimentais produz uma correlação entre uma variável de personalidade (como extroversão) e uma variável comportamental (como um ato de ajudar). Correlações variam de -1 a +1. Uma correlação 0 significa que a variável individual é literalmente inútil para prever o resultado comportamental; uma correlação 1 significa que a variável individual é um preditor positivo perfeito; uma correlação -1 significa que a variável individual é um preditor negativo perfeito. Correlações reais tendem a estar entre -.3 e +.3. A quantidade de variação explicada por uma dada variável preditora é o quadrado da correlação entre esta variável e o comportamento em questão. Assim, por exemplo, se extroversão está correlacionada com o comportamento de ajudar em .25, então extroversão explica 6.25% da variação no comportamento de ajudar. Embora esta seja apenas uma das medidas de poder explicativo, e seja bastante simplista, variáveis de personalidade não parecem melhores sob outras medidas, como o d de Cohen, n2 e parcial-n2..

que fatores situacionais explicam aproximadamente o dobro da variação no comportamento humano do que fatores relativos aos cinco traços principais (RAUTHMANN *et al*. 2014). Evidência convergente de estudos lexicográficos e inquéritos indicam que existem pelo menos cinco dimensões das situações que são confiáveis para a predição de pensamento, sentimento e comportamento: (1) valência negativa, (2) adversidade, (3) dever, (4) complexidade, (5) valência positiva (RAUTHMANN; SHERMAN, 2018).

De acordo com Doris (2002), a melhor explicação para esta falta de consistência intersituacional é que a grande maioria das pessoas tem traços locais, e não globais. Elas não são honestas, corajosas ou gananciosas, mas podem ser honestas-quando-de-bom-humor, corajosas-enquanto-navegam-com-amigos-durante-uma-tempestade, gananciosas-a-menos-que-estejam-sendo-observadas-por-colegas-da-paróquia. Em contraste, Christian Miller (2013, 2014) pensa que a evidência é melhor explicada por uma teoria dos traços globais mistos, tal como a disposição de alguém de (dentre outras coisas) ajudar porque isto melhora seu humor. Estes traços são globais, no sentido de que explicam e predizem o comportamento em diferentes situações (mantendo-se os outros fatores inalterados, alguém com tal disposição tipicamente ajudará ao próximo, contanto que seu humor seja mantido). Entretanto, estes traços são normativamente mistos, no sentido de que não são virtudes nem vícios. Mark Alfano (2013) vai em uma terceira direção, argumentando que atribuições de virtude e vício tendem a funcionar como profecias autorrealizáveis. Pessoas tendem a agir de acordo com os traços que são atribuídos a elas, sejam os traços virtudes menores como o asseio (MILLER; BRICKMAN; BOLEN, 1975) e a consciência ecológica (CORNELISSEN *et al*., 2006, 2007), virtudes maiores como a caridade (JENSEN; MOORE, 1977), a cooperatividade (GRUSEC *et al*., 1978) e a generosidade (GRUSEC; REDLER, 1980), ou vícios como a competitividade predatória (GRUSEC *et al*., 1978). Na perspectiva de Alfano, quando pessoas agem de acordo com uma virtude, elas muitas vezes o fazem não porque possuem o traço em questão, e sim porque elas pensam que possuem o traço, ou porque sabem que outras pessoas pensam que elas o possuem. Ele chama tais simulações de caráter moral de "virtudes factícias", e chega a sugerir que a noção de uma virtude deve ser revisada para incluir expectativas reflexivas e sociais.[75]

[75] Merritt (2000) foi o primeiro a sugerir que poderíamos lidar com a crítica situacionista escoando parte da responsabilidade pela virtude para o meio social em um modo análogo a este.

Pode parecer que as críticas que motivam estas novas abordagens da virtude erram o alvo. Afinal, os eticistas da virtude não precisam se comprometer (e muitas vezes não se comprometem) com a afirmação de que quase todo mundo é virtuoso. Ao contrário, muitos argumentam que a virtude é o objetivo normativo do desenvolvimento moral, e que a maioria das pessoas falha de várias maneiras em alcançar este objetivo. Tais perspectivas podem ver como um *non sequitur* ao argumento que parte do fato de que a maioria das disposições das pessoas não são virtudes para a rejeição da versão ortodoxa da ética da virtude. Mas os críticos empiricamente informados da ética da virtude não param por aí. Todos eles têm perspectivas positivas sobre quais tipos de disposições as pessoas possuem **ao invés de** virtudes. Sustenta-se que estas disposições são estruturalmente tão diferentes das virtudes (como tradicionalmente entendidas) que pode ser psicologicamente pouco realista tratar virtudes (tradicionais) como um ideal regulador. O que importa, então, é a largura do abismo entre o descritivo e o normativo, entre (a estrutura das) disposições que a maioria das pessoas possui e (a estrutura das) disposições que contam como virtudes.

Existem três respostas dominantes a esta crítica. Alguns eticistas da virtude (KUPPERMAN, 2009) concederam que a virtude é extremamente rara, mas argumentaram que, ainda assim, ela pode ser um ideal regulador útil. Outros (HURKA, 2006; MERRITT, 2000) tentaram enfraquecer o conceito de virtude, de modo a permitir que mais indivíduos, ou pelo menos mais comportamentos, sejam contados como virtuosos. Ainda outros (KAMTEKAR, 2004; RUSSELL, 2009; SNOW, 2010; SREENIVASAN, 2002) desafiaram a evidência situacionista ou a interpretação dela. Embora ainda não esteja claro se estas defesas são bem-sucedidas, o enfrentamento do desafio situacionista levou tanto defensores como críticos da ética da virtude a desenvolver perspectivas mais sutis e empiricamente informadas.[76]

---

[76] Pode-se esperar que a reflexão filosófica sobre ética promova o comportamento moral. Eric Schwitzgebel começou recentemente a investigar se eticistas profissionais se comportam moralmente melhor do que seus colegas filósofos não eticistas. De acordo com ele, sob a maioria dos padrões, os dois grupos são indistinguíveis (SCHWITZGEBEL, 2009; SCHWITZGEBEL; RUST, 2010; SCHWITZGEBEL et al., 2011).

3.2 Bem-estar, Felicidade e a Vida Boa

Os filósofos sempre se interessaram pelo que faz uma vida humana correr bem, mas as últimas décadas testemunharam um crescimento dramático na pesquisa psicológica e filosófica sobre felicidade, bem-estar e o que contribui para uma vida boa. É importante distinguir entre uma "vida boa" no sentido de uma vida que corre bem para aquele que a vive, e uma vida **moralmente** boa, pois há muito que filósofos discutem se uma pessoa moralmente má poderia desfrutar de uma vida boa. O estudo empírico do bem-estar se concentra primariamente em vidas que são boas no primeiro sentido: boa para a pessoa que a vive. Em segundo lugar, precisamos distinguir entre uma vida hedonisticamente boa e uma vida globalmente boa. A vida hedonisticamente boa é aquela que o sujeito experiencia como prazerosa. Uma vida globalmente boa pode não conter tanto prazer, mas ainda assim ser boa por outras razões, como aquilo ela alcança. Pode ocorrer de, após investigação, decidirmos que uma vida globalmente boa tem de ser uma vida hedonisticamente boa, mas os dois conceitos são distintos.

Com estas distinções em mente, podemos ver a relevância da evidência experimental para investigações nesta área. A primeira e talvez mais óbvia relevância é que experimentos podem investigar nossas intuições sobre o que constitui uma vida boa, de modo a fornecer insights sobre os conceitos comuns de felicidade, bem-estar e florescimento. Com este propósito, Phillips, Nyholm e Liao (2014) investigaram intuições sobre a relação entre moralidade e felicidade. Seus resultados sugerem que a concepção comum de felicidade envolve tanto componentes descritivos como componentes normativos: em particular, julgamos que pessoas são felizes se estão experienciando sentimentos positivos que elas devem experienciar. Assim, para usar um exemplo deles, um médico nazista que adquire sentimentos positivos ao conduzir seus experimentos não é feliz. Em contraste, uma enfermeira que adquire sentimentos positivos ao ajudar crianças doentes é feliz, embora uma enfermeira que se sinta mal pelas mesmas ações não seja feliz.

Outro conjunto de descobertas experimentais relevantes para esta questão envolve o experimento de pensamento da máquina de experiências, de Nozick (1974, p. 44-45). Em resposta à afirmação hedonista de que o prazer é o único bem intrínseco, Nozick nos pede que consideremos o seguinte:

> Suponha que exista uma máquina de experiências que lhe daria qualquer experiência que desejasse.

> Super neuropsicólogos poderiam estimular seu cérebro, de modo que você pensaria e sentiria que está escrevendo um grande romance, fazendo um amigo ou lendo um livro interessante. O tempo todo você estaria flutuando em um tanque, com eletrodos acoplados ao seu cérebro. Será que você deveria se conectar a esta máquina por toda a vida, pré-programando suas experiências de vida? […] Evidentemente, enquanto você estiver no tanque você não saberá que está lá; você pensará que tudo está realmente acontecendo [...] Você se conectaria?

Nozick argumenta que nossa resposta à máquina de experiências revela que bem-estar não é algo puramente subjetivo: "Ao imaginar uma máquina de experiências e então perceber que não a usaríamos, aprendemos que algo além da experiência é importante para nós". DeBrigard (2010) relata ter descoberto que as intuições dos sujeitos sobre a experiência eram diferentes quando eles eram informados de que já estavam na máquina. Neste caso, eles não escolheriam se desconectar. DeBrigard explica isto em termos de um viés em favor do *status quo*. Weijers (2014) vai mais longe ao perguntar aos sujeitos sobre o cenário original de Nozick e pedir que justifiquem suas respostas. Ele descobriu que muitas justificações implicavam um tipo de resistência imaginativa ao cenário, ou citavam fatores irrelevantes. Ele também descobriu que os respondentes tinham maior propensão em dizer que "uma vida conectada" seria melhor quando eles estavam escolhendo para outra pessoa, e não para eles próprios.

Um segundo tipo de estudo envolve a investigação das causas e correlatos da felicidade, bem-estar e satisfação com a vida. Estes experimentos examinam as condições sob as quais pessoas relatam emoções positivas ou satisfação com a vida, bem como as condições sob as quais eles julgam que suas vidas estão indo bem. Este tipo de pesquisa é distinto do primeiro. Afinal, o fato de que uma pessoa relata uma experiência como sendo prazerosa não necessariamente nos informa sobre se ela julgaria que aquela experiência é prazerosa para outra pessoa; pode haver assimetrias entre avaliações de primeira e terceira pessoa. Ademais, embora este tipo de pesquisa possa nos contar como, de uma perspectiva de primeira pessoa, vários elementos contribuem para o bem-estar, isto não resolve a questão do conceito de uma vida boa. Pode ocorrer de eu julgar que minha vida está indo bem e, no entanto, falhar em perceber que meu juízo se deve ao fato de eu estar

de muito bom humor, enquanto a verdade é que não estou alcançando qualquer de meus objetivos. Se eu for perguntado acerca de outra pessoa cuja situação é similar, talvez não faça o mesmo juízo. Qual destes juízos melhor representa nosso conceito de bem-estar é uma questão complicada, e também normativa, já que evidência experimental por si só pode não a resolver. Como se vê, experimentos revelaram uma série de fatores que influenciam nossos relatos e avaliações de nosso próprio prazer e bem-estar. Discutiremos duas áreas de pesquisa em particular: relatos de prazeres e dores, e juízos de satisfação com a vida.

Primeiro, no domínio da avaliação hedônica existem claras divergências entre as somas agregadas de prazeres e dores momentâneas e memórias *ex post* dos prazeres e dores. Por exemplo, o nível de dor recordado de uma colonoscopia é bem predito pela média entre o pior nível momentâneo de dor e o nível final de dor. Ademais, a duração do procedimento não tem qualquer efeito mensurável sobre a classificação *ex post* da dor (REDELMEIER; KAHNEMAN, 1996). Isto significa que os resumos posteriores feitos por pessoas de suas experiências hedônicas não são inegrais simples dos teores hedônicos momentâneos ao longo do tempo. Se a colonoscopia fosse funcionalmente completada após o minuto 5, mas arbitrariamente prolongada por outros 5 minutos, de modo que o nível final de dor fosse menor ao fim do minuto 10 do que ao fim do minuto 5, o paciente avaliaria retrospectivamente a experiência como menos dolorosa. Isto dificulta explicações do bem-estar em termos de prazer (por exemplo, o hedonismo de Bentham (1789/1961)), na medida em que levanta a questão de se o prazer medido é o prazer experienciado no momento ou retrospectivamente. Se, ao longo do próprio ato de agregar prazer, nós mudamos a forma como o avaliamos, então temos uma complicação para o hedonismo e algumas versões de utilitarismo. Dado que bem-estar alegadamente é uma noção normativa capaz de guiar tanto ações individuais como políticas sociais, descobertas como estas também colocam em questão o que exatamente devemos buscar a fim de maximizá-lo: momentos prazerosos ou a avaliação retrospectiva global do prazer.

Tais descobertas levaram alguns filósofos a procurar por alternativas ao hedonismo, na esperança de estabelecer uma base empiricamente mais estável para o bem-estar: em particular, a ideia de que bem-estar consiste em satisfação com a vida. O psicólogo mais proeminente no campo é Ed Diener[77] , cuja Escala

[77] Veja, por exemplo, Diener, Scollon e Lucas (2003).

de Satisfação com a Vida solicita que os participantes concordem ou discordem com afirmações como “Eu estou satisfeito com minha vida” e “Se pudesse viver minha vida novamente, eu mudaria quase nada”. Estas questões parecem atingir características mais estáveis e significativas da vida do que avaliações hedônicas em termos de prazer e dor, e pode-se esperar que as respostas a elas sejam mais consistentes e robustas. Entretanto, isto levanta dois problemas. O primeiro é que as respostas dos participantes aos questionários de satisfação com a vida podem não ser relatos precisos de atitudes permanentes. Por exemplo, Fox e Kahneman (1992) mostraram que, especialmente com respeito a domínios pessoais que as pessoas parecem valorar (amigos e vida amorosa), a resposta dos participantes é predita por comparação social, e não por fatores intrapessoais recentes. Uma pessoa que acabou de perder um amigo, mas ainda pensa em si como alguém que tem mais amigos que seus colegas, tenderá a relatar maior satisfação com a vida do que uma pessoa que acabou de ganhar um amigo, mas ainda pensa em si como alguém que tem menos amigos que seus colegas.

Inquéritos sobre satisfação com a vida também parecem estar sujeitos a efeitos de ordem. Por exemplo, se um participante é questionado sobre sua satisfação global com a vida e, em seguida, sobre sua vida amorosa, a correlação entre estas duas questões tende a ser próxima de zero; mas se o participante é primeiramente questionado sobre sua vida amorosa, a correlação tende a ser alta e positiva (STRACK; MARTIN; SCHWARZ, 1988).[78] Isto sugere que juízos de satisfação com a vida podem ser indevidamente influenciados pelos assuntos ou questões considerados logo antes do juízo ser feito. Outro exemplo é o frequentemente citado estudo de Schwarz e Clore, de 1983, no qual os autores relataram uma correlação entre o tempo e a satisfação com a vida, quando os sujeitos eram questionados primeiramente sobre o tempo, mas não quando a questão acerca do tempo vinha depois da questão sobre a satisfação com a vida. Deve-se notar, contudo, que existe uma controversa significativa sobre este último exemplo. Alguns artigos recentes afirmam encontrar uma correlação entre os dois (p. ex. CONNOLLY, 2013), enquanto outros afirmam que não existe qualquer evidência para um efeito do clima

[78] Para uma resposta crítica, veja Schimmack e Oishi (2005), que argumentam que informação cronicamente acessível é um preditor muito melhor de respostas de satisfação com a vida do que informação temporalmente acessível, como a quantos encontros uma pessoa foi na semana passada.

sobre juízos de satisfação com a vida (*vide* LUCAS; LAWLESS, 2013).

Descobertas como estas acima levaram alguns pesquisadores (*vide* HAYBRON, 2008) a argumentar que juízos de satisfação com a vida são demasiado arbitrários para satisfazer os requisitos para uma teoria do bem-estar. Em resposta, Tiberius e Plakias (2011) argumentam a favor de uma teoria idealizada da satisfação com a vida, que eles chamam "**satisfação com a vida baseada no valor**". Eles sugerem que muito da instabilidade que assola os estudos descritos acima pode ser removida se pedirmos aos sujeitos que, ao considerar sua satisfação com a vida, se atentem aos domínios que eles mais valoram. Eles apoiam esta afirmação por meio de uma pesquisa que mostra que juízos feitos após sujeitos serem pré-ativados para pensar sobre seus valores exibem altos níveis de estabilidade de reteste (SCHIMMACK; OISHI, 2005).

3.3. Emoção

Como muito da discussão anterior revela, a relação entre emoção e juízo moral é um dos interesses centrais da filosofia moral tradicional e da filosofia moral experimental. Nossa discussão deste tópico se concentrará em dois tipos de pesquisa: o papel da emoção no raciocínio moral, e o papel de uma emoção específica – o nojo – nos juízos morais.

Um dos debates diz respeito ao problema de se o raciocínio inflamado, guiado por emoções, é necessariamente melhor ou pior do que o raciocínio que se basea apenas no pensamento frio, mais refletido. Por vezes, emprega-se a terminologia de sistema 1/sistema 2 de pensamento, proposta por Kahneman, para falar desta distinção. Todavia, a terminologia não é perfeita. Isto porque os termos de Kahneman dizem respeito a uma distinção entre um sistema de juízos rápido, automático e inconsciente (sistema 1) e um processo de tomada de decisão lento, trabalhoso e deliberativo (sistema 2). Como vimos acima, esta não é uma distinção entre emoção e razão, pois juízos baseados em regras podem ser automáticos e inconscientes, enquanto juízos emocionais podem ser trabalhosos e conscientes. O debate entre Mikhail e Haidt diz respeito à medida em que nossos juízos morais são explicados por emoções, ao invés de regras; os argumentos de Singer e Greene contra o deontologismo se apoiam na afirmação de que juízos baseados na emoção são menos justificados do que suas contrapartes utilitaristas.

Uma razão para pensar que juízos morais essencialmente envolvem um componente emocional é que eles tendem a nos motivar. **Internalismo** é a perspectiva de que existe uma conexão necessária entre juízo moral e motivação. Esta perspectiva se contrasta com o **externalismo**. O último não nega que juízos morais normalmente são motivadores, mas nega que eles são necessariamente motivadores, e afirma que o elo entre juízo e motivação é somente contingente. Dado que emoções são intrinsecamente motivacionais, mostrar que juízos morais consistem pelo menos parcialmente de emoções justificaria o internalismo. Um caminho para esta conclusão envolveu pesquisar as intuições das pessoas sobre o juízo moral. Nichols (2002, p. 289) perguntou a sujeitos se um agente que "não tem qualquer reação emocional ao ferir outras pessoas", mas afirma saber que ferir outros é errado, realmente entende que ferir outros é errado. Ele descobriu que a maioria dos sujeitos de fato atribuiu conhecimento em tais casos, sugerindo que o conceito comum de juízo moral é externalista, tese também apoiada por Strandberg e Björklund (2013). Uma fonte adicional de evidência vem dos psicopatas e pacientes com lesões cerebrais traumáticas. Ambos servem como evidência de funcionamento moral deficiente, mas o modo como esta evidência é tomada depende da perspectiva que adotamos: enquanto externalistas (NICHOLS, 2002; ROSKIES, 2003) afirmam que a existência de psicopatas fornece um contraexemplo à tese de que o juízo moral é motivador (pois psicopatas não possuem empatia ou aversão a causar dano a outros), internalistas (SMITH, 1994; MAIBOM, 2005; KENNETT; FINE, 2008) argumentam que psicopatas, na verdade, não fazem juízos morais completos. O psicólogo Robert Hare (1993, p. 129) cita um pesquisador que diz que eles "conhecem as palavras, mas não a música". Psicopatas também exibem outras deficiências cognitivas e emocionais, como evidenciado por suas precárias habilidades de tomar decisões em outras áreas. Talvez isto signifique que eles não devam ser tomados como evidência contra o internalismo. Uma razão para pensar que juízos morais devem não envolver emoção é que por vezes emoções parecem levar a juízos distorcidos, que em outras circunstâncias nos pareceriam injustificados. Uma emoção em particular é muitas vezes mencionada no contexto desta discussão: o nojo. Esta emoção, que parece ser exclusiva de animais humanos e surge em um período relativamente tardio do desenvolvimento (a partir dos 5-8 anos de idade), envolve uma característica expressão facial boquiaberta, uma tendência a se afastar do objeto do nojo, uma leve redução na temperatura corporal e frequência cardíaca, e uma sensação de náusea e necessidade de se limpar. Além disto, o sujeito enojado é tipicamente motivado a evitar ou mesmo expurgar o objeto ofensivo, ele o

experiencia como contagioso e repugnante, fica mais atento a outros objetos nojentos nas proximidades, e inclina-se a também tratar como nojenta qualquer coisa com a qual o objeto tenha contato (físico ou metafórico). Esta última característica é muitas vezes chamada de "potência de contaminação" e é um dos aspectos que torna o nojo tão potente e, de acordo com seus críticos, tão problemáticos. A reação de nojo pode ser difícil de reprimir, é facilmente reconhecida, e induz, empaticamente, o nojo naqueles que a reconhecem.[79] Existem certas coisas das quais quase todos os adultos normais sentem nojo (fezes, cadáveres em decomposição, comida estragada, aranhas, larvas, deformidades físicas extremas). Para além destes objetos centrais de nojo, contudo, também existe considerável variação intercultural e interpessoal, onde a emoção se estende para escolhas alimentares, comportamentos sexuais, membros de fora do grupo, e violações de normas sociais. Muitos estudos alegaram mostrar que o nojo é implicado em juízos morais mais severos (SCHNALL; HAIDT; CLORE, 2008), citando experimentos nos quais sujeitos que preencheram questionários em salas fedorentas ou sujas avaliaram transgressões morais de forma mais severa. Outros foram mais longe e argumentaram que o nojo pode causar ou fazer parte de um juízo moral (HAIDT, 2001; WHEATLEY; HAIDT, 2005). Críticos argumentam que, se isto for verdade, devemos ser cautelosos com estes juízos. Isto porque o nojo tem um passado conturbado em vários sentidos. Primeiro, ele é historicamente (e também atualmente) associado com racismo, sexismo, homofobia e xenofobia; a linguagem do nojo é muitas vezes usada em campanhas de discriminação e até genocídio. Segundo, a história evolutiva do nojo nos fornece razão para desconfiança. Kelly (2011) argumenta que as manifestações corporais universais de nojo evoluíram para ajudar humanos a evitar a ingestão de toxinas e outras substâncias prejudiciais, enquanto o sentido mais cognitivo e simbólico de ofensividade e contaminação associado ao nojo evoluiu para ajudar humanos a evitar doenças e parasitas. Posteriormente, este sistema foi recrutado para um propósito inteiramente distinto: ajudar a demarcar as fronteiras entre endogrupo e exogrupo, e assim motivar a cooperação com os membros do grupo , a punição de desertores e a exclusão dos membros de fora do grupo.

---

[79] Veja May (2014) para uma crítica destas descobertas, e Kelly (2011, em particular o capítulo 1) para um exame extenso da literatura.

Se a explicação do nojo proposta por Kelly estiver no caminho correto, então parece que temos vários resultados morais importantes. Uma consequência, argumenta ele, é o "ceticismo quanto ao nojo" (2011, p. 139), de acordo com o qual a combinação entre o gatilho sensível do nojo e sua trajetória balística significa que ele é particularmente propenso a falsos positivos incorrigíveis, que envolvem sentimentos injustificados de contaminação e até desumanização. Consequentemente, "o fato de que algo é nojento nem remotamente é um indicador confiável de jogo sujo moral", é, ao contrário, "irrelevante para a justificação moral" (2011, p. 139).

É importante notar que as considerações céticas levantadas por Kelly são específicas ao nojo e sua história evolutiva particular, de modo que elas não pretendem minar o papel de todas as emoções no juízo moral. Ainda assim, se Kelly está correta e se o nojo é implicado em muitos juízos morais, talvez tenhamos uma razão para o ceticismo com respeito a muitos de nossos juízos. Plakias (2011, 2017) argumenta contra a primeira metade desta antecedente, alegando que Kelly e outros "céticos do nojo" estão errados ao afirmar que os propósitos do nojo moral e do nojo físico são totalmente distintos. Ela sugere que às vezes o nojo é uma resposta adequada a violações morais, que nos protege do contágio social. May (2014) argumenta conta a segunda metade, alegando que o papel do nojo no juízo moral tem sido significativamente exagerado. O máximo que temos é evidência de que, em certos casos, o nojo pode intensificar levemente juízos morais que já foram formados. Entretanto, trabalho empírico mais recente indica que, embora o nojo incidental tenha pouco efeito sobre a dureza dos juízos morais, a sensibilidade disposicional ao nojo tem este efeito (LANDY; GOODWIN, 2015).

4. Metaética e Filosofia Moral Experimental

A metaética dá um passo atrás na teorização moral para perguntar sobre a natureza e função da moralidade. Enquanto a ética de primeira ordem procura explicar o que devemos fazer, a metaética procura explicar o status destas teorias de primeira ordem: Quais são seus compromissos metafísicos, e quais evidências temos a favor e contra eles? Qual epistemologia melhor caracteriza nossas práticas morais? Qual é a semântica correta da linguagem moral? Estas questões podem não parecer obviamente empíricas. Entretanto, visto que a metaética tenta explicar a semântica, a epistemologia e a ontologia da moral, seu objetivo parcial é captar ou explicar o que fazemos quando nos envolvemos em conversas, juízos e avaliações

morais. Na medida em que a metaética vê a si própria como caracterizando nossa prática cotidiana da moralidade, ela é sensível a dados empíricos sobre esta prática. O fato de que uma teoria afirma que estamos enganados ou em erro generalizado sobre o significado da linguagem moral, ou que nos falta justificação para nossas crenças morais centrais, é encarado como um golpe na teoria. Por exemplo, o relativismo é muitas vezes criticado por exigir que abandonemos a crença (supostamente) difundida de que afirmações morais dizem respeito a questões de fato, objetivas, que são verdadeiras ou falsas independentemente de nossas crenças e atitudes. Já vimos vários modos de dados experimentais serem relevantes para teorias acerca de razões morais (o debate entre internalistas e externalistas) e para a epistemologia do juízo moral (o debate sobre o papel das intuições). Nesta seção, examinaremos contribuições experimentais para o debate sobre o realismo moral, sobre os argumentos acerca do desacordo moral e sobre a linguagem moral.

4.1. Metaética de Senso Comum e Realismo Moral

Muito da metaética contemporânea se apoia em suposições sobre a natureza do pensamento, discurso e prática moral cotidianos; a metaética tende a ver seu projeto como essencialmente conservador. Por exemplo, Michael Smith (1994) escreve que o primeiro passo da teorização metaética é "identificar características que se manifestam na prática moral cotidiana", e o segundo é "dar sentido a uma prática que tem estas características". Esta suposição teve um grande impacto no debate entre realistas e antirrealistas, com os primeiros afirmando captar melhor a natureza do discurso moral cotidiano: "Nós começamos como cognitivistas e realistas (tácitos) sobre ética", afirma Brink (1989), e então "somos levados a alguma forma de antirrealismo (supondo que sejamos) apenas porque passamos a considerar que os compromissos do realista moral são insustentáveis... O realismo moral deveria ser nosso ponto de partida metaético, e deveríamos abandoná-lo apenas se ele de fato envolver compromissos metafísicos e epistemológicos inaceitáveis".

Entretanto, o trabalho experimental lançou dúvidas sobre esta afirmação, começando com Darley and Goodwin (2008), e continuando com James Beebe (2014) e outros (WRIGHT *et al.,* 2013; CAMPBELL; KUMAR, 2012; GOODWIN; DARLEY, 2010; e SARKISSIAN *et al.,* 2011). Goodwin e Darley pediram que sujeitos classificassem seu acordo com afirmações extraídas dos domínios da moral, ética e estética, e perguntaram aos sujeitos se eles concordam com a afirmação (por

exemplo, “antes do terceiro mês de gravidez, o aborto, por qualquer razão (da mãe) que seja, é moralmente permissível”), se eles pensam que ela representa um **fato** ou uma **opinião ou atitude**, e se, caso alguém discordasse deles quanto à afirmação, pelo menos um dos discordantes teria de estar enganado. (Na próxima seção, falaremos mais sobre o uso que os autores fazem do desacordo como um indicador de realismo). Em geral, sujeitos classificaram afirmações morais como menos factuais do que afirmações obviamente factuais (por exemplo, “a Terra é o centro do universo”), porém mais factuais do que afirmações sobre questões de gosto ou etiqueta. O que impressiona nestas descobertas não é apenas que as pessoas não são inequivocamente realistas, mas que elas aparentemente tratam questões morais de forma variada: algumas são tratadas como questões de fato, outras como questões de opinião. Este padrão foi replicado por vários estudos, e persiste mesmo quando é permitido que os próprios sujeitos determinem quais assuntos são atribuídos ao domínio moral. Isto sugere que os sujeitos não pensam que afirmações morais são uniformemente objetivas.

Uma questão que permanece é como explicar esta variação: será que as pessoas estão sendo inconsistentes? Será que estão remetendo a algo além da questão moral – talvez ao grau de consenso que existe sobre uma questão? Pesquisas futuras podem nos ajudar a jogar alguma luz sobre estas questões. Todavia, mesmo que as questões sejam respondidas afirmativamente, a tese dos realistas de que eles captam a moralidade de senso comum é colocada em questão. Isto porque estes experimentos sugerem que ou as intuições de senso comum não são uniformemente realistas ou estão misturadas (e, consequentemente, são pouco adequadas para desempenhar um papel na teorização metaética). Outra resposta a estas descobertas é que elas revelam um tipo de pluralismo metaético de senso comum. Por exemplo, Michael Gill (2008, 2009) e Walter Sinnott-Armstrong (2009) sugeriram um tratamento não cognitivista para algumas áreas da moralidade e cognitivista para outras. Esta resposta é pouco motivada por dados empíricos, posto que estes dados mostram, no máximo, que há variação em nossos juízos sobre a medida em que afirmações morais descrevem fatos independentes da mente, mas não que a própria semântica de tais juízos varia. Tais dados também não resolvem a questão de se nossos juízos morais são crenças ou outro estado conativo: embora, como discutido acima, dados experimentais possam revelar um papel para a emoção no juízo moral, eles não mostram que em tais casos a crença está ausente. Na seção 5, abaixo, discutiremos em maior detalhe limitações da filosofia moral experimental. Por enquanto, parece que é melhor interpretar os dados sobre o

realismo de senso comum como minando uma das fontes mais comumente citadas de evidência para o realismo moral, nomeadamente, que ele capta nosso discurso e prática moral cotidianos melhor do que seus concorrentes. Uma questão adicional, aberta à exploração, é se uma perspectiva antirrealista pode se sair melhor do que as concorrentes realistas em captar e explicar nossas intuições de senso comum.

4.2. Desacordo Moral

Como vimos no início deste verbete, alguns dos primeiros pensamentos empiricamente informados em filosofia moral diziam respeito ao desacordo moral. Brandt e Ladd conduziram investigações minuciosas dos códigos morais de outros grupos, e alguns filósofos morais contemporâneos argumentaram a favor de uma atenção permanente aos detalhes empíricos do desacordo moral. Em parte, isto se deve à formulação de Mackie (1977, p. 36-37) do que ele chamou de o "argumento da relatividade" em termos de uma inferência para a melhor explicação: a melhor explicação da natureza e persistência da diversidade moral, argumenta Mackie, é que nossos juízos morais representam "modos de vida", ao invés de "percepções, na maior parte seriamente inadequadas e severamente distorcidas, de valores morais objetivos". A tendência dos realistas foi responder ao argumento apontando para outras explicações possíveis do desacordo, que são consistentes com a objetividade, tais como equívocos sobre fatos não morais, irracionalidade, e falhas de imparcialidade e imaginação (veja por exemplo Brink (1984), Sturgeon (1988), Smith (1994) e Shafer-Landau (2003); para uma apresentação geral e análise das respostas realistas, veja Loeb (1998)).

Aqui é útil lembrar da advertência de Brandt de que "informações isoladas sobre diferenças interculturais" não serão suficientes para resolver o debate. O argumento do desacordo se apoia tanto na existência do desacordo moral como na melhor explicação para ele. Por esta razão, a avaliação das perspectivas do argumento requer atenção a detalhes empíricos de desacordos morais reais, ao invés de conjecturas sobre os resultados de desacordos morais possíveis. Por exemplo, Doris e Plakias (2008) discutem várias instâncias de desacordo moral intercultural, e avaliam as perspectivas de aplicar o que chamam de "explicações dissipadoras" a estes casos. A explicação dissipadora é aquela que explica o desacordo em termos de uma diferença não moral, ou de uma falha epistêmica da parte de um ou mais dos discordantes, de modo a mostrar que não se tratava

realmente de um desacordo moral. Seu argumento e os experimentos que eles citam são discutidos em detalhe no verbete sobre **psicologia moral empírica**, de modo que não iremos discuti-los em detalhe aqui; por enquanto, notaremos que é difícil avaliar estas explicações apenas por meio de métodos de inquérito. Esta é a razão de ser particularmente útil olhar para o registro antropológico, como faz Oliver Curry *et al.* (no prelo) em uma recente publicação pioneira. Curry levanta a hipótese de que a moralidade abrange soluções estáveis e cooperativas a problemas recorrentes que podem ser modelados como jogos de soma não zero. Existe uma variedade de tais problemas recorrentes, e a teoria dos jogos estabeleceu um conjunto de ferramentas analíticas para diagnosticar e resolvê-los. Curry e seus colaboradores mostraram que estas soluções sempre são vistas como ou moralmente boas ou moralmente neutras em sociedades ao longo de vários estágios de desenvolvimento, desde pequenos grupos de caçadores-coletores até democracias industrializadas. Esta pesquisa sugere que filósofos podem ter exagerado muito a extensão do desacordo moral que realmente existe.

4.3 Linguagem Moral

Uma razão pela qual o desacordo é uma medida útil de objetividade é que a impossibilidade de desacordo na ausência de erro é considerada uma característica de problemas que dizem respeito às questões de fato, objetivas. Por outras palavras, até podemos conceder a possibilidade de desacordo sem erro com respeito a assuntos que envolvem o quão gostosa é uma comida ou quais os méritos de uma equipe esportiva. Mas quando discordamos sobre assuntos morais, pensamos que existe uma resposta correta e que, portanto, pelo menos um dos discordantes está errado. Para uma discussão da analogia entre comida e moralidade, consulte Loeb (2003). Na medida em que pessoas julgam que um desacordo não envolve erro, podemos pensar que estão indicando um tipo de antirrealismo sobre a questão em disputa. Esta é mais uma maneira pela qual o desacordo pode ser relevante para o debate sobre realismo/antirrealismo. Pode a evidência experimental envolvendo o desacordo nos contar qualquer coisa sobre a semântica da linguagem moral?

Uma razão para pensar que ela pode é que a própria ideia de desacordo sem erro parece incoerente a alguns. Nos casos em que um indivíduo pensa (ou diz) que sorvete de pistache é delicioso e outro pensa (ou diz) que é nojento, nós entendemos os dois envolvidos como expressando suas próprias preferências

pessoais. Para explicar a aparente ausência de erro na disputa, nós reinterpretamos seus pensamentos ou proferimentos como expressando algo como "sorvete de pistache é delicioso (nojento) para mim". Dado que as duas partes não se contradizem realmente, não precisamos dizer que uma delas está errada. Portanto, uma vez que a ausência de erro parece implicar que não existe qualquer conteúdo único sobre o qual as partes discordam (e sobre o qual uma delas pode estar errada), ela alegadamente é um ponto a favor do antirrealismo moral.

Em contraste, realistas afirmam que sua perspectiva é mais capaz de captar nossa intuição de que há um desacordo genuíno em jogo quando uma pessoa diz que roubar é errado e outra diz que roubar não é errado. Realistas argumentam que, nestes casos, temos a intuição de que realmente há um conflito, e esta intuição pode ser captada somente por uma teoria que assume que a linguagem moral envolve atribuição de propriedades a ações ou coisas, ao invés de meros relatos ou expressões de atitudes de um falante.

Nós já vimos alguns dados relevantes para este debate em nossa discussão do realismo de senso comum, acima. A alegação aqui não diz respeito ao realismo propriamente dito, mas se a evidência experimental pode fornecer *insight* sobre a semântica da linguagem moral, posto que nossas intuições sobre a existência de uma ausência genuína de erro podem nos informar sobre se existe um único conteúdo compartilhado por dois proferimentos – uma única proposição que é afirmada por uma das partes e negada pela outra – ou sobre se os proferimentos dos dois envolvidos contém relativizações implícitas. No mínimo, esta foi a suposição que norteou grande parte do debate sobre desacordo e sua implicação para a semântica moral. Em um trabalho recente, contudo, John Khoo e Joshua Knobe (2016) lançaram dúvidas sobre esta suposição. Seus experimentos indicam que sujeitos não veem o desacordo como algo que requer conteúdo exclusivo – uma única proposição que uma parte aceita e a outra rejeita. Esta pesquisa sugere que intuições sobre o desacordo podem não estar tão diretamente ligadas à semântica moral quanto filósofos pensaram: juízos de que dois indivíduos genuinamente discordam sobre alguma afirmação não necessariamente implicam um único conteúdo compartilhado por dois falantes. Embora este trabalho ainda esteja nos estágios iniciais, ele revela as complicações em extrair uma semântica a partir do desacordo.

Um desafio adicional para tentativas de investigar experimentalmente a semântica moral é o debate dos antirrealistas entre cognitivismo e não cognitivismo. De acordo com o cognitivismo, frases morais comuns são melhor compreendidas como afirmações factuais, que expressam proposições que atribuem propriedades

morais (tais como correção ou bondade) a coisas (tais como ações ou caráter). Não cognitivistas (incluindo os expressivistas, que são seus representantes contemporâneos) sustentam que a linguagem moral tem uma função fundamentalmente diferente, como, por exemplo, expressar atitudes ou emitir ordens. Embora este debate pareça propício à investigação experimental, não cognitivistas tipicamente reconhecem que sua perspectiva é reformista. Ademais, a semântica de nossos termos pode ser opaca aos usos comuns, de modo que não possamos extrair a semântica a partir de intuições e devamos investigá-la indiretamente. Por fim, o cognitivista e o não cognitivista podem estar de acordo sobre uma série de características empíricas do juízo e discurso moral; o desafio para os experimentadores é encontrar predições que confirmam uma perspectiva ao mesmo tempo em que desconfirmam a outra. Esta é uma área relativamente nova de investigação, e, embora levante alguns desafios, está aberta à exploração.

5. Críticas à Filosofia Moral Experimental

A filosofia moral experimental vai muito além do que podemos cobrir aqui, e muitas questões e áreas produtivas de pesquisa mal foram tocadas. Por exemplo, evidências experimentais são relevantes para questões morais em bioética, como eutanásia, aborto, triagem genética e intervenções placebogênicas. Do modo similar, experimentos em economia comportamental e psicologia cognitiva estão sendo empregados na apresentação de questões morais sobre políticas públicas (BICCHIERI; CHAVEZ, 2010; BICCHIERI; XIAO, 2009). Nós negligenciamos estes debates somente em razão da falta de espaço. Nas poucas palavras que nos restam, exploramos algumas potenciais críticas à filosofia experimental.

5.1. Problemas com o *Design* e Interpretação Experimental

Como um campo jovem, a filosofia experimental enfrenta vários problemas com respeito ao design e à interpretação experimental. Os problemas em questão não são intransponíveis, e também são enfrentados por áreas relacionadas, como a psicologia social, a psicologia cognitiva e a economia comportamental. Uma

questão que veio à tona recentemente é o problema da **replicação**.[80] Análise estatística não é inferência dedutiva, e o mero fato de que a análise estatística produz um resultado positivo não garante que qualquer coisa tenha sido descoberta. Tipicamente, um resultado experimental é tratado como "real" se seu valor-p é no máximo .05. Mas este valor apenas indica a probabilidade de que a observação em questão seria feita se a hipótese nula fosse verdadeira. Não se trata da probabilidade de que a hipótese nula seja falsa, dado a observação.[81] Assim, mesmo quando a análise estatística indica que a hipótese nula deve ser rejeitada, esta indicação pode ser falaciosa. Além disto, quando várias hipóteses são testadas, aumenta-se exponencialmente a chance de se rejeitar falaciosamente a hipótese nula pelo menos uma vez.

Devemos esperar outras falhas na replicação devido a um viés incorporado ao sistema de financiamento de experimentos e publicação de resultados. A proporção de falsos positivos publicados é muito maior para resultados inesperados ou não previstos do que para resultados esperados ou previstos. Dado que experimentadores hesitam em relatar resultados nulos (e são até desencorajados a fazê-lo pelos editores e pareceristas de periódicos), cujo valor-p é maior do que .05, para cada resultado inesperado publicado, pode haver qualquer número de não-resultados inesperados que não foram publicados. Deste modo, resultados inesperados

---

[80] O problema da replicação atingiu (com ou sem razão) uma proporção tão grande que John Doris, recentemente, passou a chamá-lo de "Repligate"

[81] Há também uma controversa em curso em torno do teste de significância da hipótese nula (TSHN). Em resumo, o problema é que um valor-p é uma probabilidade condicional, mas não a probabilidade condicional que se poderia esperar. Um valor-p é a probabilidade de que o resultado relevante teria sido observado dado a hipótese nula, ou seja, dado que nada de interessante está acontecendo (nenhuma correlação positiva, nenhuma correlação negativa, nenhuma interação de efeitos, e assim por diante). Por vezes, isto é investido por pesquisadores e intérpretes descuidados, que interpretam o valor-p como a probabilidade da hipótese nula dado a observação. Simbolicamente, a diferença é entre P(observação | nula) para P(nula | observação). A última probabilidade condicional, mais desejável, pode ser estimada por meio de Análise Estatística Bayesiana, embora raramente isto seja feito (existem controversas em torno da Análise Bayesiana, em particular a arbitrariedade das probabilidades iniciais). Para uma introdução a estes problemas, veja Abelson (1997), Cohen (1994) e Wagenmakers et al. (2012).

publicados são mais propensos a serem falsos positivos do que pode parecer.

A melhor maneira de dizer se um resultado tem algum valor evidencial é através da replicação do experimento – de preferência por outro grupo de pesquisa. Se um resultado não pode ser replicado, provavelmente é uma miragem. Como Daniel Kahneman famosamente apontou, estas miragens atingiram níveis preocupantes recentemente (*vide* YONG, 2012; WAGENMAKERS *et al.,* 2012). Kahneman propôs uma "ligação em cascata" de replicações, na qual nenhum resultado em psicologia seria publicado até que fosse replicado com sucesso por outro laboratório respeitável. Embora esta proposta ainda não tenha sido (e talvez nunca seja) instituída, ela colocou o problema da replicação em evidência, e um projeto relacionado decolou. O projeto de reprodutibilidade em psicologia tem o objetivo de estabelecer em que medida resultados proeminentes publicados podem ser replicados.[82] Filósofos experimentais seguiram o exemplo com seu próprio projeto de replicação (veja Outros Recursos da Internet). Recentemente, este trabalho gerou frutos animadores: um enorme esforço colaborativo de replicação sugere que aproximadamente 7 em cada 10 resultados da filosofia experimental podem ser reproduzidos, enquanto esforços similares mostram que menos da metade dos resultados em biomedicina, economia e psicologia são replicáveis (COVA *et al*., 2018; Outros Recursos da Internet).

Uma questão relacionada para o design e a interpretação experimental tem a ver com "expedições de pesca". Por vezes, refere-se pejorativamente a um experimento como uma expedição de pesca, se ele não propõe quaisquer predições específicas antes da coleta dos dados, principalmente quando várias hipóteses são testadas. Do mesmo modo, críticos argumentam que pesquisadores como Weinberg, Nichols e Stich (2008) avançaram hipóteses tão vagamente formuladas a ponto de serem praticamente infalsificáveis. Este problema é agravado quando não se propõe um guia heurístico positivo para explicar os padrões observados.

Outra preocupação é que o teste simultâneo de muitas sondagens de intuições pode levar experimentadores incautos a uma busca em vão. Suponha que um filósofo experimental conduza um experimento com duas condições: na condição experimental, participantes são pré-ativados com ideias deterministas, enquanto na condição de controle eles não são pré-ativados de qualquer modo.

---

[82] Veja o *Open Science Framework Reproducibility Project: Psychology*, disponível em: https://osf.io/ezcuj/wiki/home/. Acesso em: 25 jul. 2022.

Ele faz aos participantes vinte perguntas diferentes sobre suas intuições morais, por exemplo, se existe livre-arbítrio, se malfeitores merecem ser punidos, se a virtude é digna de recompensa, e assim por diante. Na tentativa de descobrir se a pré-ativação determinista induz mudanças nas intuições morais, ele então faz comparações pareadas de suas respostas a cada uma das questões. Deste modo, ele faz vinte comparações independentes, cada uma no nível padrão do setor, de 5%. Agora, suponha, para fins de argumentação, que não exista qualquer efeito – que todas as hipóteses nulas sejam verdadeiras. Neste caso, a probabilidade de que pelo menos um destes testes resulte em um erro de Tipo I (rejeitar a hipótese nula embora ela seja verdadeira) é de 64%. De modo mais geral, quando um experimentador realiza **n** comparações independentes ao nível de 5%, a probabilidade de pelo menos um erro de Tipo I é 1 - $.95^{n}$. Pode-se lidar com este problema por meio de vários procedimentos, mais notadamente por meio do método Tukey e do método Bonferroni[83] .

Um modo de estabelecer que um resultado é real é conferir se ele tem um tamanho de efeito respeitável. A medida mais comum do tamanho de efeito é o **d**, de Cohen, que é a razão da diferença entre as médias das condições e o desvio padrão da variável relevante. Assim, um **d** de 1.0, por exemplo, indicaria que uma manipulação afastou a média da condição experimental a um desvio padrão inteiro da média da condição de controle, um efeito enorme. Infelizmente, embora as coisas pareçam estar mudando, até recentemente os tamanhos de efeito raramente eram relatados em filosofia experimental.

---

[83] Em uma crítica recente deste tipo de pensamento estatístico falacioso, Peter Austin, Muhammad Mamdani, David Juurlink e Janet Hux (2006) descrevem argumentos estatísticos que pretendem mostrar que os signos astrológicos de pacientes canadenses estavam frequentemente correlacionados com suas patologias. Por exemplo, usando as mesmas técnicas estatísticas defendidas por muitos filósofos experimentais, seríamos levados a concluir que geminianos têm uma chance 30% maior de se tornarem alcóolatras ($p < 0.02$), escorpianos têm um risco 80% maior de desenvolver leucemia ($p < 0.05$), e mulheres de virgem sofrem 40% mais de vômito excessivo durante a gravidez ($p < 0.0.4$). Presumivelmente, estas são anomalias estatísticas, e não indicadores de riscos genuínos para a saúde.

5.2 Problemas Filosóficos

Pode-se objetar que o real problema com a filosofia moral experimental não tem a ver com a ciência, mas com sua relevância (ou, mais precisamente, irrelevância) para a filosofia moral. A objeção é que a filosofia moral não está interessada em como somos e no que fazemos, e sim em como devemos ser e no que devemos fazer. Deste modo, ela é um empreendimento normativo, imune a resultados empíricos. Ciência é difícil, mas moralidade também é; o fato de que falhamos em compreender e viver de acordo com as exigências da última não mostra um problema com nossas teorias morais, mas sim com nossas naturezas morais.

Até certo ponto, os filósofos experimentais podem concordar com esta objeção. Ninguém está sugerindo que nós extraímos diretamente a moralidade a partir de resultados de inquéritos.

Todavia, retornando ao ponto com o qual começamos, embora a moralidade seja normativa, ela também é essencialmente prática. A teoria moral é uma teoria sobre o que devemos fazer: uma teoria sobre como criaturas como nós devem se comportar e interagir com os outros. Uma moralidade completamente divorciada de nossa natureza, que exigisse ações que são impossíveis para nós, seguramente seria inaceitável. Este ponto não se aplica apenas à filosofia experimental: o utilitarismo é por vezes criticado por fazer exigências pouco realistas de imparcialidade. Ao lado da famosa proibição de se derivar um **deve** a partir de um **é**, nós encontramos a máxima de que **deve implica pode**. A exata medida em que a moralidade pode fazer exigências sem ser pouco realista é em si uma questão filosófica, de modo que a filosofia moral experimental trabalha em conjunção com a análise filosófica; ela não tem o objetivo de eliminar esta análise. É uma questão controversa como exatamente as duas áreas se relacionam, e como a evidência empírica influenciará e será relevante para debates em teoria moral no futuro. Mas é certo que a teoria moral tradicional e a filosofia moral experimental têm muito a aprender uma com a outra.

Precisamos proceder cuidadosamente aqui. Ninguém duvida que o que devemos fazer depende de como as coisas são não moralmente. Por exemplo, a afirmação moral de que determinado homem merece ser punido pressupõe o fato não moral de que ele cometeu um crime. Portanto, não deveria ser surpresa que evidência experimental pode ser de tal modo relevante para a moralidade. Mas ainda é uma questão em aberto se evidência experimental é relevante para a descoberta dos princípios morais fundamentais – aqueles que devem ser aplicados de uma forma ou de outra, a depender de como o mundo é não moralmente.

Outra versão deste argumento diz que os princípios filosóficos morais fundamentais são, quando verdadeiros, necessariamente verdadeiros, e que a pesquisa empírica pode no máximo estabelecer verdades contingentes.[84] Mas se as teorias morais fundamentais são necessárias, então são necessárias **para criaturas como nós**. E algo que a investigação empírica pode fazer é nos ajudar a estabelecer quais tipos de criaturas nós somos. A imaginação precisa de material para trabalhar. Quando alguém se senta em sua poltrona e imagina um cenário hipotético, faz uma série de suposições sobre como as pessoas são, como processos psicológicos funcionam, e assim por diante. Estas suposições podem ser empiricamente bem ou mal informadas. É difícil ver como alguém poderia duvidar que ser bem informado é melhor. Do mesmo modo, é difícil ver como alguém poderia duvidar da relevância da evidência experimental para uma melhor fundamentação de nossas suposições empíricas, e, no presente caso, de nossas suposições relevantes para a filosofia moral. Como exatamente a evidência experimental – e, de forma mais geral, empírica – é relevante, e quão relevante ela é, são questões atualmente muito debatidas.

Referência bibliográfica

ABELSON, R. On the surprising longevity of flogged horses: Why there is a case for the significance test. **Psychological Science**, 8 (1): 12–15, 1997.

ADAMS, F.; STEADMAN, A. Intentional action in ordinary language: Core concept or pragmatic understanding? **Analysis**, 64: 173–181, 2004.

ADAMS, F.; STEADMAN, A. Folk concepts, surveys, and intentional action, in ed. LUMER, C. **Intentionality, deliberation, and autonomy**: The action-theoretic basis of practical philosophy. Aldershot: Ashgate, p. 17–33, 2007.

ALFANO, M. **Character as Moral Fiction.** Cambridge: Cambridge University Press, 2013.

ALFANO, M. **Moral Psychology**: An Introduction. London: Polity, 2016.

ALFANO, M.; BEEBE, J.; ROBINSON, B. The centrality of belief and reflection in Knobe-effect cases. **The Monist**, 95(2): 264–289, 2012.

[84] Estamos em dívida com Chris Heathwood aqui.

ALICKE, M. Blaming badly. **Journal of Cognition and Culture**, 8: 179–186, 2008.

ANNAS, J. **Intelligent Virtue.** Oxford: Oxford University Press, 2011.

APSLER, R. Effects of embarrassment on behavior toward others. **Journal of Personality and Social Psychology,** 32: 145–153, 1975.

AUSTIN, P.; MAMDANI, M.; JUURLINK, D.; HUX, J. Testing multiple statistical hypotheses resulted in spurious associations: A study of astrological signs and health. **Journal of Clinical Epidemiology**, 59 (9): 964–969, 2006.

AYER, A. J. **Language Truth, and Logic.** London: Gollancz,1936.

BANERJEE, K.; HUEBNER, B.; HAUSER, M. Intuitive moral judgments are robust across demographic variation in gender, education, politics, and religion: A large-scale web-based study. **Journal of Cognition and Culture**, 10 (1/2): 253–281, 2010.

BARON, R. The sweet smell of … helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls. **Personality and Social Psychology Bulletin**, 23: 498–503,1997.

BARON, R.; KENNY, D. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. **Journal of Personality and Social Psychology,** 51: 1173–82,1986.

BARON, R.A.; THOMLEY, J. A whiff of reality: Positive affect as a potential mediator of the effects of pleasant fragrances on task performance and helping. Environment and Behavior, 26: 766–784, 1994.

BEEBE, J. R. A Knobe effect for belief ascriptions. **Review of Philosophy and Psychology**, 4 (2): 235–258, 2013.

BEEBE, J. R. How different kinds of disagreement impact folk metaethical judgments, in ed. WRIGHT, J. C.; SARKISSIAN, H. **Advances in Experimental Moral Psychology**: Affect, Character, and Commitment. London: Continuum, p. 167–187, 2014.

BEEBE, J. R.; BUCKWALTER, W. The epistemic side-effect effect. **Mind & Language**, 25: 474–498, 2010.

BEEBE, J.R.; JENSEN, M. Surprising connections between knowledge and action: The robustness of the epistemic side-effect effect. **Philosophical Psychology**, 25 (5): 689–715, 2012.

BENTHAM, J. **An Introduction to the Principles of Morals and Legislation**. Garden City: Doubleday, 1961. Originalmente publicado em 1789.

BICCHIERI, C.; CHAVEZ, A. Behaving as expected: Public information and fairness norms. **Journal of Behavioral Decision Making**, 23(2): 161–178, 2010.

BICCHIERI, C. AND XIAO, E. Do the right thing, but only if others do so. **Journal of Behavioral Decision Making**, 22: 191–209.2009.

BLAKE, P.; MCAULIFFE, K.; WARNEKEN, F. The developmental origins of fairness: The knowledge-behavior gap. **Trends in Cognitive Sciences**, 18 (11): 559–561, 2014.

BLOOM, P. **Just Babies**: The Origins of Good and Evil. New York: Crown, 2013.

BONNEFON, J. F.; SHARRIFF, A.; RAHWAN, I. The social dilemma of autonomous vehicles. **Science**, 352 (6293): 1573–1576, 2016.

BRANDT, R. **Hopi Ethics**: A Theoretical Analysis. Chicago: University of Chicago Press,1954.

BRANDT, R. **A Theory of the Right and the Good**. Prometheus Books,1998.

BRINK, D. Moral realism and the sceptical arguments from disagreement and queerness. Australasian Journal of Philosophy, 62 (2): 111–125, 1984.

BUCKWALTER, W.; STICH, S. Gender and philosophical intuition, in ed. KNOBE; NICHOLS. **Experimental Philosophy** (Volume 2). Oxford: Oxford University Press, 2014.

CAMPBELL, R.; KUMAR, V. Moral reasoning on the ground. Ethics, 122 (2): 273–312, 2012.

CAPPELEN, H. Philosophy Without Intuitions. Oxford: Oxford University Press, 2012.

CASE, T.; REPACHOLI, B.; STEVENSON, R. My baby doesn't smell as bad as yours: The plasticity of disgust. Evolution and Human Behavior, 27 (5): 357–365, 2006.

CHOMSKY, N. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press, 1965.

COHEN, J. The Earth is round ($p < .05$). American Psychologist, 49 (12): 997–1003, 1994.

CORNELISSEN, G.; DEWITTE, S.; WARLOP, L.; *et al.* Free bumper stickers for a better future: The long term effect of the labeling technique. **Advances in Consumer Research**, 33: 284–285, 2006.

CURRY, O.; MULLINS, D.; WHITEHOUSE, H. Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. **Current Anthropology** (no prelo).

CUSHMAN, F.; GREENE, J. Finding faults: How moral dilemmas illuminate cognitive structure. **Social Neuroscience**, 7 (3): 269–279, 2012.

CUSHMAN, F.; YOUNG, L. The psychology of dilemmas and the philosophy of morality. **Ethical Theory and Moral Practice**, 12 (1): 9–24, 2009.

CUSHMAN, F.; YOUNG, L. Patterns of judgment derive from nonmoral psychological representations. **Cognitive Science**, 35(6): 1052–1075, 2011.

DARLEY, J.; BATSON, C. D. From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. **Journal of Personality and Social Psychology**, 27: 100–108, 1973.

DE BRIGARD, F. If you like it, does it matter if it's real? **Philosophical Psychology**, 23(1): 43–57, 2010.

DEUTSCH, M. Arguments, intuitions, and philosophical cases: A note on a metaphilosophical dialectic. **Philosophical Forum**, 47 (3–4): 297–307, 2016.

DIENER, E.; SCOLLON, C.; LUCAS, R. The evolving concept of subjective wellbeing: The multifaceted nature of happiness. **Advances in Cell Aging and Gerontology**, 15: 187–219, 2003.

DIENER, E.; EMMONS, R.; LARSEN, R. *et al.* The satisfaction with life scale. **Journal of Personality Assessment**, 49 (1): 71–5, 2010.

DORIS, J. Persons, situations, and virtue ethics. **Noûs**, 32(4): 504–540, 1998.

DORIS, J. **Lack of Character**: Personality and Moral Behavior. Cambridge: Cambridge University Press, 2002.

DORIS, J.; PLAKIAS, A. How to argue about disagreement: Evaluative diversity and moral realism, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, p. 303–332, 2008b.

DORIS, J.; PLAKIAS, A. How to Find a Disagreement: Philosophical Diversity and Moral Realism, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, p.345–54,2008b.

DORIS, John; STICH, Stephen. Moral Psychology: Empirical Approaches. The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), 2012.

DWYER, S. Moral competence, in ed. MURASUGI, K.; STAINTON, R. **Philosophy and Linguistics**. Boulder, CO: Westview Press, p. 169–190, 1999.

DWYER, S. Moral dumbfounding and the linguistic analogy: Implications for the study of moral judgment. **Mind and Language**, 24: 274–96, 2009.

FLANAGAN, O. **Varieties of Moral Personality**: Ethics and Psychological Realism. Cambridge: Harvard University Press, 1983.

FOOT, P. **Virtues and Vices and Other Essays in Moral Philosophy**. Berkeley, CA: University of California Press; Oxford: Blackwell, 1978.

FOX, C.; KAHNEMAN, D. Correlations, causes and heuristics in surveys of life satisfaction. **Social Indicators Research**, 27: 221–34, 1992.

FUNDER, D.; OZER, D. Behavior as a function of the situation. **Journal of Personality and Social Psychology**, 44: 107–112, 1983.

GILL, M. Metaethical Variability, Incoherence, and Error, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, p. 387–402, 2008b.

ARMSTRONG, W. Indeterminacy and variability in meta-ethics. **Philosophical Studies**, 145 (2): 215–234, 2009.

GOODWIN, G.; DARLEY, M. The psychology of metaethics: Exploring objectivism. **Cognition**, 106 (3): 1339–1366, 2008.

GOODWIN, G.; DARLEY, J. The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy. **Review of Philosophy and Psychology**, 1 (2): 161–188, 2010.

GREENE, J. The secret joke of Kant's soul, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Neuroscience of Morality: Emotion, Brain Disorders, and Development (Volume 3). Cambridge, MA: MIT Press, p. 35–80, 2008c.

GREENE, J. Reflection and reasoning in moral judgment. **Cognitive Science**, 36 (1): 163–177, 2012.

GREENE, J.; MORELLI, S.; LOWENBERG, K.; *et al.* Cognitive load selectively interferes with utilitarian moral judgment. **Cognition**, 107 (3): 1144–1154, 2008.

GREENE, J.; NYSTROM, L.; ENGELL, A.; *et al.* The neural bases of cognitive conflict and control in moral judgment. **Neuron**, 44: 389–400, 2004.

GREENE, J.; SOMMERVEILLE, R.; NYSTROM, L.; *et al*. An fMRI investigation of emotional engagement in moral judgment. **Science**, 293: 2105–8, 2001.

GRUSEC, J.; REDLER, E. Attribution, reinforcement, and altruism: A developmental analysis. **Developmental Psychology**, 16 (5): 525–534,1980.

GRUSEC, J.; KUCZYNSKI, L.; RUSHTON, J.; *et al*. "Modeling, direct instruction, and attributions: Effects on altruism. **Developmental Psychology**, 14: 51–57, 1978.

HAIDT, J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. **Psychological Review**, 108 (4): 814–834, 2001.

HAIDT, J. **The Righteous Mind**: Why Good People are Divided by Politics and Religion. New York: Pantheon, 2012.

HAIDT, J.; BJÖRKLUND, F. Social intuitionists answer six questions about moral psychology, in SINNOTT-ARMSTRONG, W. **Moral Psychology**: The

Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, 181–218. 2008b.

HALLEN, B. **Knowledge, Belief, and Witchcraft**: Analytic Experiments in African Philosophy. Oxford: Oxford University Press, 1986.

HALLEN, B. **The Good, the Bad, and the Beautiful**: Discourse about Values in Yoruba Culture, Bloomington: Indiana University Press, 2000.

HARE, R. **Without Conscience**: The Disturbing World of the Psychopaths Among Us. New York: Guilford Press, 1993.

HARMAN, G. Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. **Proceedings of the Aristotelian Society** (New Series), 119: 316–331, 1999.

HARMAN, G. The nonexistence of character traits. **Proceedings of the Aristotelian Society**, 100: 223–226, 2000a.

HARMAN, G. **Explaining Value and Other Essays in Moral Philosophy**. New York: Oxford University Press, 2000b.

HARMAN, G. **No character or personality**. Business Ethics Quarterly, 13 (1): 87–94, 2003.

HARMAN, G. Using a linguistic analogy to study morality, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Evolution of Morality: Adaptations and Innateness (Volume 1). Cambridge, MA: MIT Press, p. 345–352, 2008a.

HAUSER, M. **Moral Minds**: How Nature Designed a Universal Sense of Right and Wrong. New York: Ecco Press/Harper Collins, 2006.

HAUSER, M.; CHOMSKY, N.; FITCH, W. The faculty of language: What is it, who has it, and how does it evolve. **Science**, 298: 1569–1579, 2002.

HAUSER, M.; YOUNG, L.; CUSHMAN, F. Reviving Rawls's linguistic analogy: Operative principles and the causal structure of moral actions, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, p. 107–144. 2008b.

HAYBRON, D. **The Pursuit of Unhappiness**. Oxford: Oxford University Press, 2008.

HEATHWOOD, C. The problem of defective desires. **Australasian Journal of Philosophy**, 83 (4): 487–504, 2005.

HURKA, T. Virtuous act, virtuous dispositions. **Analysis**, 66 (289): 69–76, 2006.

INBAR, Y.; PIZARRO; KNOBE, J.; *et al.* Disgust sensitivity predicts intuitive disapproval of gays. Emotion, 9 (3): 435–443, 2009.

ISEN, A.; LEVIN, P. The effect of feeling good on helping: Cookies and kindness. **Journal of Personality and Social Psychology**, 21: 384–88, 1972.

JACKSON, F. **From Metaphysics to Ethics**. Oxford: Oxford University Press, 1998.

JENSEN, A.; MOORE, S. The effect of attribute statements on cooperativeness and competitiveness in school-age boys. **Child Development**, 48: 305–307, 1977.

KAHNEMAN, D.; DIENER, E.; SCHWARTZ, N. (eds). **Wellbeing**: The Foundations of Hedonic Psychology. New York: Russell Sage, 2003.

KAHNEMAN, D.; DIENER, E.; SCHWARTZ, N. (eds). **Wellbeing**: The Foundations of Hedonic Psychology. New York: Russell Sage, 2003.

KAHNEMAN, D. **Thinking, Fast and Slow.** New York: Farrar, Straus, & Giroux, 2011.

KAMTEKAR, R. Situationism and virtue ethics on the content of our character. Ethics, 114 (3): 458–491, 2004.

KELLY, D. **Yuck!** The Nature and Moral Significance of Disgust. Cambridge: MIT Press, 2011.

KENNETT, J. Do psychopaths really threaten moral rationalism? **Philosophical Explorations**, 9 (1): 69–82, 2006.

KHOO, J.; KNOBE, J. Moral disagreement and moral semantics. **Noûs**, 52 (1): 109–143, 2018.

KLINE, R. B. **Principles and Practice of Structural Equation Modeling**. New York: Guilford Press, 2005.

KLUCKHOHN, C. **Mirror for Man.** New York: McGraw Hill,1959.

KNOBE, J. Intentional action and side effects in ordinary language. **Analysis**, 63 (3): 190–194, 2003.

KNOBE, J. Folk psychology and folk morality: Response to critics. J**ournal of Theoretical and Philosophical Psychology**, 24: 270–279, 2004a.

KNOBE, J. Intention, intentional action and moral considerations. **Analysis**, 2: 181–187. 2004b.

KNOBE, J. The concept of intentional action: A case study in the uses of folk psychology. **Philosophical Studies**, 130(2): 203–231, 2006.

KNOBE, J. Reason explanation in folk psychology. **Midwest Studies in Philosophy**, 31: 90–107, 2007.

KNOBE, J. Cause and norm. **Journal of Philosophy**, 106 (11): 587–612, 2009.

KNOBE, J. Person as scientist, person as morality. **Behavioral and Brain Sciences**, 33: 315–329, 2010.

KNOBE, J. Is morality relative? Depends on your personality. **The Philosopher's Magazine**, 52: 66–71, 2011.

KNOBE, J. Experimental philosophy is cognitive science, in ed. SYTSMA, J.; BUCKWALTER, W. **A Companion to Experimental Philosophy**. Oxford: Blackwell, 2016.

KNOBE, J.; MENDLOW, G. The good, the bad and the blameworthy: Understanding the role of evaluative reasoning in folk psychology. **Journal of Theoretical and Philosophical Psychology**, 24: 252–258, 2004.

KOHLBERG, L. The Psychology of Moral Development: The Nature and Validity of Moral Stages. San Francisco: Harper and Row, 1984.

KUPPERMAN, J. Virtue in virtue ethics. **Journal of Ethics**, 13 (2–3): 243–255, 2009.

LADD, J. **The Structure of a Moral Code**: A Philosophical Analysis of Ethical Discourse Applied to the Ethics of the Navaho Indians. Cambridge, MA: Harvard University Press, 1957.

LANDY, J.; GOODWIN, G. Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. **Perspectives on Psychological Science**, 10 (4): 518–536, 2015.

LAM, B. Are Cantonese speakers really descriptivists? Revisiting cross-cultural semantics. Cognition, 115: 320–332, 2010.

LATANÉ, B.; DARLEY, J. Group inhibition of bystander intervention in emergencies. **Journal of Personality and Social Psychology**, 10: 215–221, 1968.

LATANÉ, B.; DARLEY, J. **The Unresponsive Bystander**: Why Doesn't He Help? New York: Appleton-Century-Crofts, 1970.

LATANÉ, B.; NIDA, S. Ten years of research on group size and helping. **Psychological Bulletin**, 89: 308–324, 1981.

LATANÉ, B.; RODIN, J. A lady in distress: inhibiting effects of friends and strangers on bystander intervention. **Journal of Experimental Psychology**, 5: 189–202, 1969.

LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, 140: 1–55, 1932.

LOEB, D. Moral realism and the argument from disagreement. **Philosophical Studies**, 90 (3): 281–303, 1998.

LOEB, D. Gastronomic value realism: A cautionary tale. **Journal of Theoretical and Philosophical Psychology**, 21 (1): 30–49.2003.

LUCAS, R. Adaptation and the set-point model of subjective wellbeing: Does happiness change after major life events? **Current Directions in Psychological Science**, 16 (2): 75–9, 2007.

LUCAS, R.; CLARK, C.; GEORGELLIS, Y.; *et al*. Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. **Journal of Personality and Social Psychology**, 84 (3): 527–539, 2003.

LUETGE, C.; RUSCH, H.; UHL, M. (eds.). **Experimental Ethics**: Towards an Empirical Moral Philosophy. London: Palgrave, 2014.

MACHERY, E.; MALLON, R.; NICHOLS, S.; *et al*. Semantics, cross-cultural style. **Cognition**, 92 (3): B1–B12, 2004.

MACINTYRE, A. **A Short History of Ethics**: A History of Moral Philosophy from the Homeric Age to the Twentieth Century. South Bend, IN: University of Notre Dame Press, 1998.

MACINTYRE, A. **After Virtue**: A Study in Moral Theory. South Bend, IN: University of Notre Dame Press,1984.

MACKIE, J. L. **Ethics**: Inventing Right and Wrong. New York: Penguin, 1977.

MAIBOM, H. L. Moral unreason: The case of psychopathy. **Mind and Language**, 20 (2): 237–57, 2005.

MATTHEWS, K. E.; CANNON, L. K. Environmental noise level as a determinant of helping behavior. **Journal of Personality and Social Psychology**, 32: 571–577, 1975.

MATTHEWS, K. E.; CANNON, L. K. Environmental noise level as a determinant of helping behavior. **Journal of Personality and Social Psychology**, 32: 571–577, 1975.

MAY, J. Does disgust influence moral judgment? **Australasian Journal of Philosophy**, 92 (1): 125–141, 2014.

MCGILVRAY, J. **The Cambridge Companion to Chomsky**. Cambridge: Cambridge University Press, 2005.

MERRITT, M. Virtue ethics and situationist personality psychology. **Ethical Theory and Moral Practice**, 3 (4): 365–383, 2000.

MIKHAIL, J. Chomsky and moral philosophy, in ed. MCGILVRAY, J. **The Cambridge Companion to Chomsky**. Cambridge: Cambridge University Press, 235–53, 2005.

MIKHAIL, J. Universal moral grammar: Theory, evidence, and the future. **Trends in Cognitive Sciences**, 11: 143–152, 2007.

MIKHAIL, J. The poverty of the moral stimulus, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Evolution of Morality: Adaptations and Innateness (Volume 1). Cambridge, MA: MIT Press, p. 353–360, 2008a.

MIKHAIL, J. Moral grammar and intuitive jurisprudence: A formal model of unconscious moral and legal language, in ed. ROSS, B. H.; BARTELS, D. M.; BAUMAN, C. W.; *et al*. **Psychology of Learning and Motivation** (Volume 50: Moral Judgment and Decision Making). New York: Academic Press, 2009.

MIKHAIL, J. **Elements of Moral Cognition**: Rawls's Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment. Cambridge: Cambridge University Press, 2011.

MILLER, C. **Moral Character**: An Empirical Theory. Oxford: Oxford University Press, 2013.

MILLER, C. **Character and Moral Psychology**. Oxford: Oxford University Press, 2014.

MILLER, R.; BRICKMAN, P.; BOLEN, D. Attribution versus persuasion as a means for modifying behavior. **Journal of Personality and Social Psychology**, 31 (3): 430–441, 1975.

MISCHEL, W. **Personality and Assessment**. New York: Wiley, 1968.

MOODY-ADAMS, M. M. **Fieldwork in Familiar Places**: Morality, Culture, and Philosophy. Cambridge, MA: Harvard University Press, 1997.

NADELHOFFER, T. On praise, side effects, and folk ascriptions of intentionality. **Journal of Theoretical and Philosophical Psychology,** 24: 196–213, 2004.

NADELHOFFER, T. Bad acts, blameworthy agents, and intentional actions: Some problems for jury impartiality. **Philosophical Explorations**, 9: 203–220, 2006.

NAESS, A. Truth as Conceived by Those Who are Not Professional Philosophers. Oslo: Jacob Dybwad, 1938.

NICHOLS, S. On the genealogy of norms: A case for the role of emotion in cultural evolution. **Philosophy of Science**, 69: 234–255, 2002.

NICHOLS, S. **Sentimental Rules**. Oxford: Oxford University Press, 2004.

NICHOLS, S.; KUMAR, S.; LOPEZ, T.; *et al*. Rational learners and moral rules. **Mind and Language**, 31 (5): 530–554, 2016.

NOZICK, R. **Anarchy, State, and Utopia**. Basic Books,1974.

NUSSBAUM, M. **The Fragility of Goodness**: Luck and Ethics in Greek Tragedy and Philosophy. Cambridge: Cambridge University Press, 2001.

PEREBOOM, D. **Living Without Free Will**. New York: Cambridge University Press, 2001.

PETTIT, D.; KNOBE, J. The pervasive impact of moral judgment. **Mind and Language**, 24 (5): 586–604, 2009.

PHELAN, M.; SARKISSIAN, H. Is the 'trade-off hypothesis' worth trading for? **Mind and Language**, 24: 164–180, 2009.

PHILLIPS, J.; MISENHEIMER, L.; KNOBE, J. The ordinary concept of happiness (and others like it). **Emotion Review**, 71: 929–937, 2011.

PHILLIPS, J.; NYHOLM, S.; LIAO, S. **The good in happiness**. Oxford Studies in Experimental Philosophy (Volume 1), Oxford: Oxford University Press (no prelo).

PIAGET, J. **The Moral Judgment of the Child**. New York: Free Press, 1965.

PLAKIAS, A. The response model of moral disgust. **Synthese,** first online 3 June 2017. doi:10.1007/s11229-017-1455-3, 2017.

PLAKIAS, A. The good and the gross. **Ethical Theory and Moral Practice**, 16 (2): 261–78, 2013.

PRINZ, J. **The Emotional Construction of Morals**. Oxford, Oxford University Press, 2007.

PRINZ, J. Resisting the linguistic analogy: A commentary on Hauser, Young, and Cushman, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, 157–170, 2008b.

RAILTON, P. The affective dog and its rational tale: Intuition and attunement. **Ethics**, 124 (4): 813–859, 2014.

RAUTHMANN, J.; SHERMAN, R. The description of situations: Towards replicable domains of psychological situation characteristics. **Journal of Personality and Social Psychology**, 114 (3): 482–488, 2018.

RAWLS, J. **A Theory of Justice**. Cambridge, MA: Harvard University Press, 1971.

REDELMEIER, D. & KAHNEMAN, D. Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. **Pain**, 1: 3–8, 1996.

REGAN, J. Guilt, perceived injustice, and altruistic behavior. **Journal of Personality and Social**

Psychology, 18: 124–132,1971.

ROBINSON, B.; STEY, P.; ALFANO, M. Virtue and vice attributions in the business context: An experimental investigation. Journal of Business Ethics, 113 (4): 649–661, 2013.

ROEDDER, E. & HARMAN, G. Linguistics and moral theory, in ed. DORIS, J. **The Moral Psychology Handbook**, 273–296. Oxford University Press, 2010.

RÖNNOW-RASUMUSSEN, T. **Personal Value**. Oxford: Oxford University Press, 2011.

ROSKIES, A. Are ethical judgments intrinsically motivational? Lessons from 'acquired sociopathy'. **Philosophical Psychology**, 16 (1): 51–66, 2003.

RUSSELL, D. **Practical Intelligence and the Virtues**. Oxford: Oxford University Press, 2009.

SARKISSIAN, H.; PARK, J.; TIEN, D.; *et al*. Folk moral relativism. **Mind and Language**, 26 (4): 482–505, 2011.

SAYRE-MCCORD, Geoffrey (ed.). **Essays on Moral Realism**. Ithaca: Cornell University Press, 1988.

SAYRE-MCCORD, Geoffrey. Moral Semantics and Empirical Inquiry, in ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, p. 403–412, 2008b.

SCAIFE, R.; WEBBER, J. Intentional side-effects of action. **Journal of Moral Philosophy**, 10: 179–203, 2013.

SCHIMMACK, U.; OISHI, S. The influence of chronically and temporarily accessible information on life satisfaction judgments. **Journal of Personality and Social Psychology**, 89 (3): 395–406, 2005.

SCHNALL, S.; HAIDT, J.; CLORE. Disgust as embodied moral judgment. **Personality and Social Psychology Bulletin**, 34 (8): 1096–1109, 2008.

SCHWARTZ, S.; GOTTLIEB, A. Bystander anonymity and reactions to emergencies. **Journal of Personality and Social Psychology**, 39: 418–430, 1991.

SCHWITZGEBEL, E. Do ethicists steal more books? **Philosophical Psychology**, 22 (6): 711–725, 2009.

SCHWITZGEBEL, E. & CUSHMAN, F. Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. **Mind and Language**, 27 (2): 135–153, 2012.

SCHWITZGEBEL, E.; RUST, J. Do ethicists and political philosophers vote more often than other professors? **Review of Philosophy and Psychology**, 1 (2): 189–199, 2010.

SCHWITZGEBEL, E.; RUST, J.; HUANG, L.; *et al*. Ethicists' courtesy at philosophy conferences. **Philosophical Psychology**, 25 (3): 331–340, 2011.

SHAFER–LANDAU, R. **Moral Realism**: A Defence. Oxford: Oxford University Press, 2003.

SINGER, P. Ethics and Intuitions. **The Journal of Ethics**, 9 (3–4): 331–352, 2005.

SINGER, P.; DE LAZARI-RADAK, K. **From the Point of View of the Universe**: Sidgwick and Contemporary Ethics. Oxford: Oxford University Press (no prelo).

SINNOTT-ARMSTRONG, W. Mixed-up meta-ethics. **Philosophical Issues**, 19 (1): 235–56, 2009.

SINNOTT-ARMSTRONG, W. (ed.). **Moral Psychology**: The Evolution of Morality: Adaptations and Innateness (Volume 1). Cambridge, MA: MIT Press, 2008a.

SINNOTT-ARMSTRONG, W. (ed.). **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, 2008b.

SINNOTT-ARMSTRONG, W. (ed.). **Moral Psychology**: The Neuroscience of Morality: Emotion, Brain Disorders, and Development (Volume 3). Cambridge, MA: MIT Press, 2008c.

SINNOTT-ARMSTRONG, W. 2008d. Framing moral intuitions. In ed. SINNOTT-ARMSTRONG, W. **Moral Psychology**: The Cognitive Science of Morality: Intuition and Diversity (Volume 2). Cambridge, MA: MIT Press, 47–76, 2008b.

SINNOTT-ARMSTRONG, W.; MALLON, R.; MCCOY, T.; *et al*. Intention, temporal order, and moral judgments. **Mind and Language**, 23 (1): 90–106, 2008.

SMITH, M. **The Moral Problem**. Oxford: Blackwell, 1994.

SNOW, N. **Virtue as Social Intelligence**: An Empirically Grounded Theory. New York: Routledge, 2010.

SREENIVASAN, G. Errors about errors: Virtue theory and trait attribution. Mind, 111 (441): 47–66, 2002.

SRIPADA, C. Punishment and the strategic structure of moral systems. Biology and Philosophy, 20 (4): 767–789, 2005.

SRIPADA, C. The deep self-model and asymmetries in folk judgments about intentional action. **Philosophical Studies,** 151 (2): 159–176, 2010.

SRIPADA, C. What makes a manipulated agent unfree? **Philosophy and Phenomenological Research**, 85 (3):

563–593, 2011. doi:10.1111/j.1933-1592.2011.00527.x

SRIPADA, C. Mental state attributions and the side-effect effect. **Journal of Experimental Psychology**, 48 (1): 232–238, 2012.

SRIPADA, C.; KONRATH, S. Telling more than we can know about intentional action. **Mind and Language**, 26(3): 353–380, 2011.

STICH, S.; WEINBERG, J. Jackson's empirical assumptions. **Philosophy and Phenomenological Research**, 62 (3): 637–643, 2001.

STRACK, F.; MARTIN, L.; SCHWARTZ, N. Priming and communication: Social determinants of information use in judgments of life satisfaction. **European Journal of Social Psychology**, 18: 429–442, 1988.

STRANDBERG, C.; BJÖRKLUND, F. Is moral internalism supported by folk intuitions? **Philosophical Psychology**, 26 (3): 319–335, 2013.

STRAWSON, P. F. Freedom and resentment. **Proceedings of the British Academy**, 48: 1–25, 1962.

STROHMINGER, N.; CALDWELL, B.; CAMERON, D. Implicit morality: A methodological survey, in ed. LUETGE, C.; RUSCH, H.; UHL, M. **Experimental Ethics**: Towards an Empirical Moral Philosophy. London: Palgrave, p. 133–156, 2014.

STURGEON, N. Moral Explanations, in ed. SAYRE-MCCORD, Geoffrey. **Essays on Moral Realism**. Ithaca: Cornell University Press, p. 229–255, 1988.

SYTSMA, J.; LIVENGOOD, J. A new perspective concerning experiments on semantic intuitions. **Australasian Journal of Philosophy**, 89 (2): 315–332, 2011.

THOMSON, J. J. A defense of abortion. **Philosophy and Public Affairs**, 1(1): 47–66, 1971.

TRIVERS, R. L. The evolution of reciprocal altruism. **Quarterly Review of Biology**, 46: 35–57, 1971.

TURIEL, E. **The Development of Social Knowledge**: Moral and Social Convention. Cambridge: Cambridge University Press,1983.

WAGENMAKERS, E.-J.; WETZELS, R.; BORSBOOM, D.; *et al*. An agenda for purely confirmatory research. **Perspectives on Psychological Science**, 7(6): 632–638, 2012.

WEIJERS, D. Nozick's experience machine is dead, long live the experience machine! **Philosophical Psychology**, 27 (4): 513–535, 2014.

WEINBERG, J.; NICHOLS, S.; STICH, S. Normativity and epistemic intuitions, in ed. KNOBE, J; NICHOLS, S. **Experimental Philosophy**. Cambridge: Cambridge University Press, 17–46, 2008.

WEYANT, J. Effects of mood states, costs, and benefits on helping. **Journal Of Personality And Social Psychology**, 36: 1169–1176, 1978.

WHEATLEY, T. AND HAIDT, J. Hypnotically induced disgust makes moral judgments more severe. **Psychological Science**, 16 (10): 780–784, 2005.

WIEGMAN, I. The evolution of retribution: Intuition undermined. **Pacific Philosophical Quarterly**, 98 (2): 490–510, 2017.

WRIGHT, J. C.; GRANDJEAN, P.; MCWHITE, C. The meta-ethical grounding of our moral beliefs: Evidence for meta-ethical pluralism. **Philosophical Psychology**, 26 (3): 336–361, 2013.

WRIGHT, J. C.; SARKISSIAN, H. (eds.). **Advances in Experimental Moral Psychology**: Affect, Character, and Commitment. London: Continuum, 2014.

YONG, E. Nobel laureate challenges psychologists to clean up their act: Social-priming research needs 'daisy chain' of replication. Nature News, 2012. Disponpivel online. doi:10.1038/nature.2012.11535

ZAGZEBSKI, L. **Exemplarist virtue theory**. Metaphilosophy, 41 (1): 41–57, 2010.

ZHONG, C. -B.; BOHNS, V.; GINO, F. Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. **Psychological Science**, 21(3): 311–314, 2010.

Agradecimentos

Os autores agradecem a James Beebe, Gunnar Björnsson, Wesley Buckwalter, Roxanne DesForges, John Doris, Gilbert Harman, Dan Haybron, Chris Heathwood, Antti Kauppinen, Daniel Kelly, Joshua Knobe, Clayton Littlejohn, Edouard Machery, Josh May, John Mikhail, Sven Nyholm, Brian Robinson, Chandra Sripada, Carissa Veliz, vários pareceristas anônimos, e aos editores desta enciclopédia pelos comentários e sugestões úteis.

Sobre os organizadores

Rodrigo Cid: Professor adjunto da Universidade Federal do Amapá, doutor em Lógica e Metafísica pela Universidade Federal do Rio de Janeiro, com pesquisa na Université Catholique de Louvain, e coordenador geral da Série Investigação Filosófica. Além de ser autor, organizador e editor de inúmeras publicações de filosofia, Prof. Cid também é um empreendedor do Metaverso e da tecnologia da *blockchain*, sendo membro fundador do The Philosophers DAO e do Metaverse DAO.

Sagid Salles: Professor da Universidade Estadual de Santa Cruz. Mestre e doutor em filosofia pelo Programa de Pós-graduação Lógica e Metafísica da UFRJ, com período sanduíche realizado na Universidade de Miami. É autor de dois livros: *Como os Nomes Nomeiam* (Editora UFPel) e *Vagueness as Arbitrariness* (Springer).

Sobre os tradutores e revisores

Cesar A. M. de Alencar: Doutor e Mestre em Filosofia pelo Programa de Pós-Graduação em Lógica e Metafísica – PPGLM/UFRJ, com pesquisa em Filosofia Antiga. Graduado em Filosofia também pela Universidade Federal do Rio de Janeiro (UFRJ). Atualmente é professor adjunto do Colegiado do Curso de Licenciatura em Filosofia – UNIFAP/Campus Santana, desenvolvendo pesquisas na área de Ética, Filosofia Política, Filosofia do Símbolo (Mente e Linguagem), Epistemologia e Crítica Literária. É coordenador do Grupo de Pesquisa Diálogo: Discurso e Ação Humana e dos projetos de extensão Tardes com Platão e Curso de Retórica em foco: Técnicas para leitura e escrita. É autor dos livros *A Filosofia de Sócrates* (Pimenta Cultural,

2021) e *Lições Socráticas* (EdUnifap, 2021). Como escritor, publicou um livro de poesia, *Poemalimpo* (2016), e um romance, *Falosofia de Mulher* (2016).

Rafael Martins: Professor do Grupo Unieduk. Doutor em Filosofia pela University of Kansas, com período como pesquisador visitante na University of Michigan, Ann Arbor. Especializado em ética normativa, metaética, várias éticas aplicadas, e razão prática.

Sérgio Farias de Souza Filho: Doutor em Filosofia pelo King's College London, mestre em Filosofia pelo Programa de Pós-Graduação Lógica e Metafísica da Universidade Federal do Rio de Janeiro e graduado em Filosofia pela Universidade Federal de Pernambuco. Pesquisador de pós-doutorado no Programa de Pós-Graduação Lógica e Metafísica da Universidade Federal do Rio de Janeiro, com Bolsa de Pós-Doutorado Nota 10 da FAPERJ. Foi pesquisador de pós-doutorado no Departamento de Filosofia da Universidade de São Paulo. Atua nas áreas de Filosofia da Mente, Filosofia da Linguagem, Metafísica, Filosofia das Ciências Cognitivas e Filosofia da Biologia. Tópicos de pesquisa: intencionalidade, naturalismo intencional, teorias teleológicas da representação mental, função biológica, seguir regras e normatividade semântica.

Sagid Salles: Professor da Universidade Estadual de Santa Cruz. Mestre e doutor em filosofia pelo Programa de Pós-graduação Lógica e Metafísica (PPGLM) da UFRJ, com período sanduíche realizado na Universidade de Miami. É autor de dois livros: *Como os Nomes Nomeiam* (Editora UFPel) e *Vagueness as Arbitrariness* (Springer).

Aluízio Couto: Doutorando e mestre em Filosofia pela UFMG e professor da rede estadual de ensino de Minas Gerais.

Vinícius Brito Barros: Licenciado em Filosofia pela Universidade Federal do Amapá - UNIFAP. Licenciando em Pedagogia na Universidade do Norte do Paraná - UNOPAR. Foi bolsista de Iniciação Científica PROBIC, na UNIFAP, entre setembro de 2021 e maio de 2022 com a temática Dialética de Zenão de Eleia. Foi bolsista de Iniciação Científica PROBIC, na UNIFAP, entre agosto de 2018 e julho de 2019 com a temática Górgias de Leontinos e a persuasão poética.

Rafael Cavalcanti de Souza: Doutorando em Filosofia pela Universidade Estadual de Campinas, mestre em Filosofia pela Universidade Estadual de Campinas e graduado em Filosofia pela Universidade Federal de Pernambuco. Desenvolve pesquisas sobre História da Filosofia da Matemática Grega e Teoria da Explicação Científica, em especial sobre a obra *Segundos Analíticos*, de Aristóteles.

L. H. Marques Segundo: Pesquisador vinculado à Associação Brasileira de Filosofia da Religião (ABFR), membro dos grupos de pesquisa Investigação Filosófica e Lógica Ladeira Abaixo (LLA). Doutor em Filosofia pela UFSC. Trabalha nas áreas de interseção entre epistemologia, metafísica e filosofia da ciência. Interessa-se, também, por abordagens evolutivas à cultura e à moralidade. Endereço eletrônico: luizhelveciosegundo@gmail.com.

Roseline Crippa: Formada pela UNAR (Centro Universitário de Araras), é professora de Literatura e Técnicas em redação do Colégio RGF (Colégio Ruth Gomes Figueira) e professora efetiva de Língua Portuguesa e Inglês (rede estadual de ensino do estado de São Paulo).

Editora
UFPel

DISSERTATIO
FILOSOFIA