



Universidade de Brasília

ROGER VALÉRIO DE VARGAS REX

O DEBATE SOBRE O NATIVISMO MORAL:

Desenvolvimento e Evolução da Moralidade

Brasília

2016

ROGER VALÉRIO DE VARGAS REX

**O DEBATE SOBRE O NATIVISMO MORAL:
Desenvolvimento e Evolução da Moralidade**

Dissertação apresentada ao Programa de Pós-graduação em Filosofia da Universidade de Brasília como requisito para a obtenção do título de mestre em Filosofia.

Orientador: Prof. Paulo César Coelho Abrantes

Brasília
2016

ROGER VALÉRIO DE VARGAS REX

O DEBATE SOBRE O NATIVISMO MORAL

Desenvolvimento e Evolução da Moralidade

Dissertação apresentada ao Programa de Pós-graduação em Filosofia da Universidade de Brasília como requisito para a obtenção do título de mestre em Filosofia.

Brasília, ___ de _____ de _____

BANCA EXAMINADORA

Prof. Paulo César Coelho Abrantes — UnB

Prof. Alexandre Araújo Costa — UnB

Prof. Cláudio Araújo Reis — UnB

Resumo

A fim de identificar o papel de estruturas psicológicas ‘inatas’ na moralidade, investigo como os seres humanos desenvolvem a capacidade de julgar comportamentos a partir de normas. Preliminarmente, procuro demonstrar que a abordagem adotada nesta dissertação não incorre na chamada falácia naturalista e, para afastar certas confusões, examino os conceitos de inato e de moralidade. Em seguida, exploro o problema sob a perspectiva de três programas de pesquisa nativistas. O primeiro deles defende a existência de uma gramática moral universal a partir de uma analogia com o modelo de princípios e parâmetros desenvolvido na Linguística. O segundo programa defende a existência de domínios morais relacionados a módulos mentais específicos, que favoreceriam o surgimento de normas com determinados conteúdos. Por fim, o terceiro programa consiste de modelos epidemiológicos, de acordo com os quais estruturas cognitivas resultantes do processo de evolução da nossa espécie provocariam vieses na origem e na transmissão das normas. Em especial, analiso como os argumentos utilizados para sustentar esses programas se saem quando confrontados com os resultados obtidos em estudos empíricos de diferentes áreas das ciências cognitivas e com argumentos de construtivistas. Quanto à teoria da gramática moral universal, argumentos como o da pobreza de estímulos e da projeção não se mostram convincentes quando aplicados ao domínio da moralidade. Estudos empíricos sobre dilemas envolvendo bondes, mencionados com frequência pelos proponentes dessa teoria, são também insuficientes para sustentar a tese de que a moralidade funciona segundo um conjunto de princípios e parâmetros. Quanto à teoria dos domínios morais, ela tem o mérito de incluir características psicológicas típicas da espécie humana, como as nossas disposições emocionais, na explicação sobre os julgamentos morais. No entanto, está comprometida com uma visão modular a respeito da mente humana que não é consiliente com os resultados de pesquisas neurocientíficas; e, além disso, não oferece uma explicação para as etapas mentais prévias ao desencadeamento das intuições morais. Com relação aos modelos epidemiológicos, o apelo a fatores de atração no processamento cognitivo oferece a hipótese mais promissora para explicar a recorrência, em diversas culturas, de normas similares. Contudo, as explicações propostas até agora são insuficientes, pois se restringem às influências de fatores psicológicos, sem explorar a sua interação com fatores de atração ecológicos. Nesse sentido, esses modelos poderiam ser complementados pela teoria da construção de nichos. Em suma, defendo que os principais programas de pesquisa nativistas não apresentam boas justificativas para sustentar a existência de estruturas cognitivas dedicadas especificamente à produção de julgamentos morais. No entanto, tendo em vista que fatores de atração psicológicos parecem canalizar o desenvolvimento da moralidade, poderíamos dizer que ela é em alguma medida inata, ao menos em um sentido ‘desenvolvimental.’

Palavras-chave: Desenvolvimento da Moralidade; Inatismo; Gramática moral universal; Domínios morais; Epidemiologia das representações; Darwinismo.

Abstract

In order to identify the role of 'innate' psychological structures in morality, I investigate how human beings develop the capacity to judge behaviors based on norms. Preliminarily, I show that the approach adopted in this dissertation does not imply the so-called naturalistic fallacy, and, to dispel confusion, I examine the concepts of innate and morality. Then, I explore the problem from the perspective of three nativist research programs. The first of them defends the existence of a universal moral grammar, by analogy with the Chomskyan model of principles and parameters in linguistics. The second program supports the existence of moral domains related to specific mental modules, which would favor the emergence of norms with certain contents. Finally, the third program consists of epidemiological models according to which cognitive structures produced by an evolutionary process would be responsible for biases in the origin and transmission of norms. In particular, I analyze how the arguments used to support these programs fare at relevant contemporary research in Cognitive Science and how well they meet arguments proposed by constructivists. With respect to the universal moral grammar theory, arguments such as the poverty of the stimulus and the projection problem are not convincing when applied to the realm of morality. Empirical studies about trolley problems, often mentioned by proponents of this theory, are also insufficient to ground the claim that morality operates according to a set of principles and parameters. Concerning the moral domains theory, it has the merit of including psychological characteristics typical of the human species, such as our emotional dispositions, in the explanation about moral judgments. Nevertheless, it is committed to a modular view of human mind that is not consistent with the results of neuroscientific research. Moreover, it does not explain the mental stages prior to the unleashing of moral intuitions. With regard to epidemiological models, the appeal to factors of attraction involved in cognitive processing offers the most promising hypothesis to explain the recurrence, in several cultures, of similar norms. However, the explanations proposed so far are still insufficient, as they are restricted to the influences of psychological factors, without exploring their interaction with ecological factors of attraction. In this sense, these models could be complemented by the niche construction theory. In short, the main nativist research programs have not presented good reasons to postulate the existence of cognitive structures specifically dedicated to the production of moral judgments. Notwithstanding, given that psychological factors of attraction seem to channel the development of morality, it is plausible to sustain that the capacity to make moral judgments is innate in some measure, at least in a developmental sense.

Keywords: Moral development; Innatism; Universal moral grammar; Moral foundations; Epidemiology of representations; Darwinism.

Sumário

Introdução	7
1 O naturalismo e as duas falácias naturalistas	13
1.1 A falácia naturalista em Moore: a indefinibilidade de ‘bom’	15
1.2 A falácia naturalista em Hume: a passagem do ser para o dever ser	18
<i>1.2.1 Primeira interpretação da lei de Hume</i>	20
<i>1.2.2 Segunda interpretação da lei de Hume</i>	24
<i>1.2.3 Relevância da advertência de Hume</i>	27
2 Nativismo moral: aspectos conceituais	31
2.1 Conceito de julgamento moral	36
<i>2.1.1 Características formais e materiais da moralidade</i>	41
<i>2.1.2 Normas morais e normas convencionais</i>	50
2.2 Conceito de inato	54
3 Moralidade e cognição	67
3.1 Cognitivismo e não cognitivismo	68
3.2 Contribuições da Psicologia Moral	73
<i>3.2.1 Emoção e julgamento moral</i>	76
<i>3.2.2 Aspectos cognitivos da moralidade</i>	82
4 Teoria da gramática moral universal	89
4.1 Argumentos da pobreza de estímulos e da projeção	91
4.2 ‘Trolley problems’ e princípio do duplo efeito	98
4.3 Desenvolvimento preestabelecido, aquisição precoce e facilidade de aprendizado ..	105
4.4 Universalidade das normas morais	111
5 Teoria dos domínios morais: Haidt	119
5.1 As origens dos domínios morais	120
5.2 Críticas à teoria dos Domínios Morais	127
<i>5.2.1 Modularidade</i>	131
<i>5.2.2 Representação mental da ação e teoria da mente</i>	139
6 Modelos epidemiológicos: Sripada e Nichols	143
6.1 Epidemiologia das representações	143
<i>6.1.1 Transmissão das representações: conservação e (re-)produção</i>	145
<i>6.1.2 Atração cultural e fatores de atração</i>	147
<i>6.1.3 Pensamento populacional: semelhanças entre Sperber e Richerson e Boyd</i>	150

6.2 Epidemiologia da moralidade	153
Conclusão	161
Referências	167

Introdução

Na tradição filosófica ocidental, discussões sobre a existência de conhecimentos inatos já ocorrem, pelo menos, desde Platão. No diálogo *Mênon*, por exemplo, ele sustenta a doutrina nativista da *anamnese*, ou seja, a ideia de que todo aprendizado é apenas uma forma de recordação de algo que já estava em nós antes mesmo de recebermos qualquer ensinamento (2005, p. 112-137). Na filosofia moderna, Descartes reviveu o nativismo, quando procurou demonstrar, em sua segunda meditação, a existência de uma lacuna entre o que percebemos pelos nossos sentidos e o que sabemos e somos capazes de compreender por meio de nossa mente (2004). No entanto, o debate nativista realmente se acirrou com o ataque de Locke, nos primeiros capítulos de seu *Ensaio sobre o Entendimento Humano*, àqueles que defendiam a existência de princípios ou ideias inatas. A posição de Locke, apesar de criticada por filósofos como Leibniz, acabou triunfando: as ciências modernas se desenvolveram à luz de um empirismo antinativista; e a doutrina das ideias inatas passou a ser vista como retrógrada, não-científica, inextricável de outras doutrinas metafísicas e teológicas que haviam caído em descrédito, e incompatível com uma abordagem naturalista sobre a natureza humana (Samet, 2008).

Entretanto esse quadro começou a sofrer grande alteração em razão dos trabalhos publicados pelo linguista Noam Chomsky, a partir dos anos 50 (1965; 1975; 1986; 2002). Chomsky defendeu que a melhor explicação para a aquisição da linguagem era a existência de um conjunto de conhecimentos inatos, chamado de ‘Gramática Universal’, que tornaria possível o aprendizado da linguagem mesmo diante da pobreza de estímulos. Após as pesquisas de Chomsky, cientistas cognitivos passaram a aderir a explicações nativistas sobre o desenvolvimento das capacidades cognitivas — rejeitando a teoria da *tabula rasa* de Locke — e a defender que conteúdos mentais estão presentes desde o nascimento. Hoje, afirmam, por exemplo, que mecanismos inatos desempenham importantes papéis nas nossas habilidades para explicar e prever o movimento de objetos físicos de tamanho médio (Spelke, 1988) e para classificar animais e plantas (Atran, 1998).

Em conformidade com essa tendência, há alguns anos filósofos e cientistas retomaram o estudo sobre o papel de estruturas inatas na moralidade. Essa preocupação com a origem do ‘senso moral’ ou da ‘faculdade moral’ não é nova na Filosofia. Quando Darwin, por exemplo, em *The Descent of Man*, se questiona sobre a origem do senso moral, ele ressalta que esse tema já fora tratado por muitos escritores de suprema habilidade (2009, p. 71). A fim de especificar quais são esses autores, ele faz referência a uma lista de 26 autores britânicos elaborada por

Alexander Bain, que incluía, entre outros, os nomes de: Hobbes, Locke, Hutcheson, Hume, Price, Smith, Austin, Whewell, J. S. Mill, Bailey, e Spencer (Bain, 2015, p. 543–724). Para a maioria dos autores dessa lista, questões empíricas referentes à existência e à origem de uma faculdade moral eram alguns dos problemas centrais — senão os problemas centrais — da filosofia moral (Mikhail, 2011, p. 187)¹.

Contudo, existem muitos filósofos morais que pensam de modo diverso. Eles acreditam que a moralidade lida com aquilo que *devemos* fazer, enquanto as ciências cognitivas apenas descrevem como as coisas *são*. Acreditam, ainda, que uma lacuna separa esses dois domínios (ser e dever ser), de forma que qualquer um que tente derivar uma conclusão acerca do que devemos fazer a partir de premissas que descrevem como as coisas são, incorre em uma espécie de falácia. Logo, pesquisas na área das ciências cognitivas seriam irrelevantes para a filosofia moral.

Entretanto, esse tipo de argumento não descreve de maneira adequada a relação entre premissas fáticas (ser) e conclusões normativas (dever ser): estudos empíricos sobre a moralidade nos fornecem uma espécie de fotografia sobre os valores vigentes; uma compreensão sobre a origem desses valores; e, conseqüentemente, ferramentas para alterá-los. Qualquer teoria normativa que pretenda ter alguma aplicação prática deveria levar em consideração esses fatores. Não há qualquer sentido, por exemplo, em recomendar condutas que as pessoas são incapazes de seguir² — dever implica poder. Além disso, se levarmos a sério a tarefa de justificar as normas morais, constataremos que, em última instância, os princípios morais que consideramos mais básicos se apoiam nas crenças que temos sobre aspectos concretos da realidade (Rosas, 2011, p. 296-7)³. No primeiro capítulo desta dissertação, aprofundo essa discussão sobre o papel das ciências na elaboração de teorias normativas e procuro demonstrar que a abordagem que adoto não implica aquilo que muitos filósofos têm chamado de ‘falácia naturalista’.

Com relação aos estudos mais recentes sobre o desenvolvimento da capacidade de julgar moralmente, chama a atenção o fato de que os conceitos de moralidade e de inato empregados não são uniformes. Em razão disso, em muitos casos, embora à primeira vista os pesquisadores procurem responder à mesma pergunta, os problemas enfrentados são distintos. Portanto, antes de abordar o debate sobre o nativismo moral propriamente dito, examinarei esses conceitos, no segundo capítulo.

¹ Ver, *e.g.*, Hume (2016) e Smith (2014).

² Para uma defesa dessa posição, ver Ruse & Wilson (1986), e Mikhail (2011, Capítulo 7).

³ Ver como Rosas (2011) relaciona explicação e justificação.

Muitas controvérsias na discussão sobre o nativismo moral surgem em razão da definição adotada para a moralidade. Responder se ela é ou não inata depende em grande medida do conceito que se adota. É possível que, dependendo da definição adotada, o nativismo moral seja falso, mas frente a outra definição seja verdadeiro. Por vezes, as controvérsias surgem em razão da falha em distinguir os elementos cognitivos e as emoções envolvidas no julgamento moral; em outros casos, pesquisadores ignoram que a nossa capacidade para fazer julgamentos morais não é uma entidade monolítica, pois envolve a utilização de diversas capacidades cognitivas, como a linguagem e a autoconsciência.

Por sua vez, o termo ‘inato’ costuma ser empregado de diversas maneiras, causando grande confusão no debate sobre o nativismo moral. Mameli e Bateson (2006), em um estudo sobre como o conceito de inato é utilizado na literatura científica, encontraram vinte e seis definições diferentes de inato! Nenhuma delas foi considerada satisfatória, pois cada uma conseguia capturar alguns dos aspectos presentes no senso comum acerca do conceito de inato, mas deixava outros de fora.

No âmbito do debate sobre o nativismo moral é comum equiparar-se a afirmação de que a moralidade é inata com a afirmação de que ela não é aprendida, ou é canalizada ambientalmente, ou então é uma adaptação darwiniana. O fato de essas definições não serem coextensivas torna o debate obscuro. Um tipo bastante comum de confusão consiste em demonstrar que uma característica possui determinada propriedade e, a partir daí, simplesmente inferir que ela também possui outras propriedades usualmente relacionadas com características inatas. A fim de evitar esse tipo de equívoco, procurarei fazer referências específicas às propriedades que estão sendo discutidas, em vez de empregar um conceito que possui tantas conotações diferentes. Assim, ficará mais claro quais são as questões empíricas que estão em discussão.

Ainda que o conceito de inato seja obscuro, isso não significa que o debate nativista em torno da moralidade seja despropositado. A percepção das ambiguidades envolvidas na utilização desse conceito pode servir para o aperfeiçoamento da abordagem adotada para tratar de debates acerca de características supostamente inatas. Há muitas questões importantes envolvidas, como veremos a seguir, que podem ser mais bem respondidas se utilizarmos conceitos claros.

O debate sobre o nativismo moral especificamente surge a partir da observação de que os seres humanos de diversas culturas classificam características comportamentais a partir de normas e valores morais. A questão que, então, se coloca e em torno da qual o debate se desenvolve é: de onde vem essa forma normativa de pensar? Essa questão pode ser respondida a

partir de pelo menos duas perspectivas. Por um lado, podemos buscar compreender como os indivíduos, ao longo de suas vidas, desenvolvem a capacidade de pensar desse modo. Por outro lado, podemos investigar a evolução dessa capacidade de valoração de condutas, ou seja, podemos averiguar quais foram os processos que levaram ao seu surgimento na espécie humana e de que forma ela vem sendo modificada ao longo do tempo.

Na presente dissertação, procurarei responder à questão acima colocada principalmente com base na primeira perspectiva mencionada, que pode ser chamada de ‘desenvolvimental.’ Ao adotar essa abordagem, a primeira hipótese que se apresenta é aquela chamada por Sripada e Stich (2007, p. 297-301) de “Pac Man.” De acordo com essa hipótese, os sistemas psicológicos que possibilitam a aquisição da capacidade para julgar moralmente não apresentam restrições ou vieses em relação aos tipos de normas a serem adotadas. As diferentes normas morais seriam itens culturais como outros quaisquer que seriam aprendidos de maneira indistinta: as crianças adquiririam todas as normas presentes em seu ambiente cultural e apenas elas. A mente humana seria como o personagem do jogo *Pac Man*, que engole tudo que está ao seu redor (Sripada & Stich, 2007, p. 297-301).

No entanto, como será demonstrado ao longo da dissertação, vários estudos psicológicos demonstram que essa posição é insustentável, tendo em vista que características psicológicas humanas resultantes do processo de evolução — muitas delas compartilhadas também por outros primatas — fazem com que certas coisas possam ser aprendidas com mais facilidade do que outras.

No outro extremo, há versões nativistas que defendem que julgamentos morais completos, incluindo o conteúdo das normas morais (proibição de assassinatos, proibição de incesto, etc.), são inatos, no sentido de não dependerem de qualquer influência cultural específica. Também não me deterei nesse tipo de nativismo forte, pois ele é minoritário mesmo entre os nativistas e tem sido rejeitado com fortes argumentos e dados empíricos. Por trás de normas como a proibição de assassinatos e a proibição de incesto — alguns dos principais candidatos a regras morais universais —, existe muito mais variabilidade do que esse tipo de nativismo sugere (Sripada, 2008, p. 324).

Analisarei versões mais fracas do nativismo, que defendem a existência de uma capacidade, em alguma medida independente de nossas experiências, para julgar moralmente ou de uma predisposição para moralizar condutas com certos conteúdos. Em especial, estudarei o problema sob a perspectiva de três tipos diferentes de programa de pesquisa.

O primeiro tipo de programa defende a existência de uma gramática moral universal a partir de uma analogia com o modelo de princípios e parâmetros desenvolvido na Linguística

por Chomsky. Conforme os defensores dessa teoria, a capacidade para a moralidade envolveria princípios inacessíveis à consciência e construções teóricas que vão além do que poderia ser explicado a partir das nossas experiências de vida. A moralidade seria o resultado desse conjunto universal de princípios e dos ajustes de parâmetros definidos culturalmente.

O segundo tipo de programa defende a existência de certos domínios morais. De acordo com esse tipo de teoria, apesar das grandes variações existentes entre as normas presentes nas diversas culturas, haveria certas intuições morais derivadas do conjunto das emoções inerentes aos seres humanos. Essas intuições, produzidas a partir de módulos mentais específicos, favoreceriam o surgimento de normas morais sobre determinados temas como reciprocidade, equidade, cooperação, condutas lesivas, *etc.*

Por fim, examinarei modelos epidemiológicos inspirados em algumas das teorias naturalistas mais consagradas acerca da evolução cultural: as de Richerson e Boyd e de Sperber. Esses modelos constituem versões mais fracas de nativismo, no sentido de que elas não requerem a existência de estruturas cognitivas dedicadas especificamente ao aprendizado moral. A dinâmica das normas morais seguiria padrões semelhantes àqueles apresentados por outros itens culturais, mas isso não significaria que elas variam de modo aleatório. Estruturas cognitivas resultantes do processo de evolução de nossa espécie provocariam vieses na origem e na transmissão das normas. Conforme essa visão, normas morais podem ser mais ou menos ‘atrativas’ em virtude do modo como elas interagem com nossas emoções, preferências e outros elementos psicológicos. Características cognitivas tipicamente humanas fariam com que certas normas fossem detectadas, armazenadas, lembradas e transmitidas mais facilmente (Sripada & Stich, 2007, p. 297-301).

Poderia também ter abordado o debate em torno do nativismo moral de outro modo, ou seja, poderia ter enfatizado o papel da evolução na conformação do comportamento e da psicologia humana. Os teóricos que adotam essa perspectiva buscam compreender como a cooperação entre humanos pode ser explicada e qual papel a moralidade pode ter desempenhado. Eles buscam identificar a maneira como a moralidade surgiu e como ela pode ter contribuído para assegurar a cooperação e aumentar a aptidão dos seres humanos. Portanto, dedicam-se essencialmente a examinar se, e em que medida, a moralidade é um produto do processo biológico de seleção natural (Joyce, 2006, p. 143; FitzPatrick, 2016).

Embora acredite que um estudo sobre *como* a moralidade surgiu e sobre qual a sua relação com a cooperação poderia auxiliar nas análises que farei nos próximos capítulos, não foi possível, nesta dissertação, enfrentar o problema do nativismo moral a partir dessa perspectiva,

tendo em vista as limitações de tempo e de espaço. Por essa razão, optei por deixar esse tipo de investigação para uma pesquisa futura.

Em suma, no primeiro capítulo, respondo a principal crítica feita por filósofos à abordagem adotada nesta dissertação. Isto é, procuro demonstrar que a investigação da moralidade a partir de uma perspectiva naturalista não envolve qualquer tipo de falácia. No segundo capítulo, esclareço alguns problemas envolvidos na utilização dos conceitos de ‘moralidade’ e de ‘inato.’ Em seguida, no terceiro capítulo, investigo quais são os papéis desempenhados pelas nossas estruturas cognitivas nos julgamentos morais. Na primeira parte desse capítulo, apresento argumentos extraídos das discussões metaéticas entre cognitivistas e não cognitivistas para demonstrar que os julgamentos morais envolvem crenças e possuem um caráter descritivo, ainda que eventualmente secundário, de forma que não podemos caracterizá-los como simples manifestações de atitudes conativas. Já na segunda parte do terceiro capítulo, analiso os estudos psicológicos mais recentes a fim de investigar a relevância das emoções e de nossas diversas capacidades intelectuais na conformação desses julgamentos. Finalmente, com base nas conclusões obtidas nesses capítulos iniciais, examino os programas de pesquisa já mencionados — teoria da gramática moral universal, teoria dos domínios morais e modelos epidemiológicos — com o objetivo de encontrar a melhor explicação sobre como os indivíduos desenvolvem a capacidade para julgar moralmente.

1 O naturalismo e as duas falácias naturalistas

A expressão falácia naturalista é utilizada para fazer referência a duas coisas bastante distintas. Essa expressão foi difundida por G. E. Moore como referência à impossibilidade de definir o ‘bom.’ No entanto, ela é frequentemente empregada para fazer referência a qualquer dedução indevida de afirmações normativas a partir de premissas puramente descritivas (Augsberg, 2008, p. 470-2). Nas seções seguintes, veremos que em qualquer um desses dois sentidos a acusação da prática dessa suposta falácia não serve como um bom argumento contra as tentativas de abordar a moralidade sob um prisma naturalista.

No entanto, antes disso, cabe aqui esclarecer qual é o tipo de naturalismo que está envolvido na presente dissertação. A fim de evitar confusões, farei uma distinção entre dois tipos de naturalismo, os quais chamarei de *naturalismo global* e de *naturalismo moral*⁴.

Na presente pesquisa, adoto como pressuposto o naturalismo global. Esse é o sentido mais geral de naturalismo, frequentemente associado a certos slogans, como a ausência de uma filosofia primeira e a continuidade entre Filosofia e Ciência. Esses slogans podem não nos dizer muito, mas ainda assim são o modo mais eficiente de sintetizar o que há de comum entre as diversas correntes naturalistas.

Uma característica comum a elas é a admiração pela Ciência e pelos seus resultados, muitas vezes combinada com uma desconfiança em relação a tentativas de elaborar uma filosofia primeira que sirva de fundamento para as ciências. Naturalistas costumam inverter essa ordem de explicação, ou seja, adotam como ponto de partida o conhecimento científico de que dispomos e começam a construir a filosofia a partir daí (Ritchie, 2008, p. 195-7).

De acordo com o naturalismo global existem apenas propriedades naturais, em algum sentido apropriado do termo natural⁵, ou seja, não existem propriedades sobrenaturais. O naturalismo global exige que apenas seja reconhecida a existência de entidades que se encaixem em uma concepção de mundo orientada pela ciência. Coisas como mesas e cadeiras superam essa exigência; fantasmas e deuses, não. No entanto, há certos objetos e propriedades misteriosas, como as propriedades morais, sobre as quais os naturalistas globais divergem, no sentido de que eles não chegam a um acordo sobre se elas existem ou não (Joyce, 2016, p. 1-2).

A ideia de continuidade entre Filosofia e Ciência, por sua vez, aponta para o aspecto metodológico do naturalismo global. A ideia de continuidade exige que as conclusões propostas em teorias filosóficas encontrem suporte ou sejam justificadas pelos resultados obtidos nos melhores estudos científicos disponíveis. Sob essa ótica, uma explicação satisfatória a respeito

⁴ Essa nomenclatura é utilizada por Joyce (2006, p. 145-6).

⁵ Sobre a distinção entre propriedades naturais e não naturais, ver Ridge (2014).

da moralidade, *e.g.*, deve ser compatível com as pesquisas existentes na área de biologia evolutiva — nossa melhor teoria a respeito de como nos tornamos o que somos (Gibbard, 2002). Uma explicação filosófica sobre a natureza e função da moralidade que seja amparada em premissas conflitantes com a teoria evolutiva não seria aceitável sob essa perspectiva. A continuidade com a ciência implica ainda que a Filosofia deve sempre que possível empregar os métodos e os estilos de explicação utilizados nas ciências — *e.g.*, identificação de causas e efeitos (Leiter, 2014).

No âmbito das discussões morais, entretanto, o termo naturalismo é muitas vezes utilizado em um sentido mais restrito para fazer referência a uma forma de realismo moral segundo a qual fatos e propriedades morais, além de existirem de modo objetivo, são fatos e propriedades naturais (Lenman, 2014). Para fazer referência a esse sentido de naturalismo, utilizarei a expressão *naturalismo moral*.

Quando deparados com propriedades misteriosas ou que parecem ser não naturais, naturalistas no sentido global frequentemente tentam reduzi-las a propriedades consideradas naturais. Uma dessas estratégias consiste em identificar as propriedades suspeitas com algum conjunto de propriedades naturais, seja por meio de definições analíticas ou de afirmações sintéticas, de modo que elas possam ser aceitas em uma ontologia naturalista. Dentre os naturalistas globais, apenas aqueles que são também naturalistas morais acreditam que a moralidade pode ser explicada por meio dessa abordagem (van Roojen, 2015).

A existência de dois significados distintos de naturalismo aponta para a existência de duas formas de *naturalizar* a moralidade identificadas por Joyce. A primeira delas seria investigar de que maneira os julgamentos e as práticas morais se encaixam em uma visão de mundo científica. Poderíamos chamar esse empreendimento de psicologia ou antropologia da moralidade. A segunda maneira seria investigar como as propriedades morais (bom, mau, certo, errado, *etc.*) se encaixam em uma visão de mundo científica. Joyce ilustra a grande diferença entre essas duas formas de naturalização por meio de uma analogia: uma coisa é buscar compreender por que as pessoas acreditam em fantasmas; outra bem diferente é tentar compreender os fantasmas em termos que sejam cientificamente respeitáveis. Embora os naturalistas morais acreditem que o segundo tipo de naturalização seja possível em relação à moralidade, muitos outros naturalistas, no sentido amplo, acreditam que propriedades morais são realmente análogas a fantasmas, ou seja, são entidades que, pela sua própria natureza, não se ajustam a uma visão de mundo orientada pela ciência (Joyce, 2016, p. 1-2).

Enquanto há bastante controvérsia sobre a possibilidade de naturalizar a moralidade no sentido pretendido pelos naturalistas morais, o prospecto de quem busca compreender a forma

como os julgamentos morais ocorrem é muito mais favorável, pois se trata de uma questão que claramente pode ser abordada por meio de métodos empíricos tradicionais. Isto é, trata-se de uma questão que pode ser respondida com o auxílio de disciplinas como a Psicologia, Sociologia, Antropologia, Neurociência, Primatologia e Economia (Joyce, 2016, p. 1-2).

Na presente pesquisa, procuro naturalizar a moralidade apenas no primeiro sentido. Isto é, busco compreender como os julgamentos morais são produzidos, mas não me dedico a prescrever condutas ou a identificar supostas propriedades morais objetivas. Como veremos a seguir, em razão de não defender o naturalismo moral, a presente pesquisa não está sujeita à acusação de praticar a falácia naturalista no sentido proposto por Moore.

1.1 A falácia naturalista de Moore: a indefinibilidade de ‘bom’

No começo do Século XX, G. E. Moore defendeu a tese de que nenhuma explicação naturalista da moralidade seria capaz de descrever com precisão o que de fato estamos pensando quando fazemos um julgamento moral. As diferentes correntes naturalistas tinham em comum o fato de identificarem propriedades morais com alguma propriedade natural, como, *e.g.*, o que é prazeroso, o que satisfaz os desejos de alguém ou o que está de acordo com a regras sociais em vigor. Moore defendeu a ideia de que os naturalistas confundiam a propriedade ‘bondade’ (*goodness*) com alguma das propriedades que as coisas boas eventualmente possuíam (Sayre-McCord, 2014).

Moore estava interessado em um problema semântico: a impossibilidade de análise de certos conceitos. Assim como não podemos, *e.g.*, definir ‘amarelo’ como “luz com um determinado comprimento de onda”, também não poderíamos definir ‘bom’ igualando-o a alguma propriedade natural. A analogia que Moore faz entre ‘amarelo’ e ‘bom’ demonstra que na visão dele a falácia naturalista não ocorria apenas quando se tentava definir propriedades relevantes para a moralidade como a ‘bondade’, mas também em qualquer tentativa de definir propriedades simples:

It is really a very simple fallacy indeed. When we say that an orange is yellow, we do not think our statement binds us to hold that ‘orange’ means nothing else than ‘yellow,’ or that nothing can be yellow but an orange. Supposing the orange is also sweet! Does that bind us to say that ‘sweet’ is exactly the same thing as ‘yellow,’ that ‘sweet’ must be defined as ‘yellow’? . . . It would be absolutely meaningless to say that oranges were yellow, unless yellow did in the end mean just ‘yellow’ and nothing else whatever—unless it was absolutely undefinable (Moore, 1922, p.14).

Moore não afirmava que a impossibilidade de definir o ‘bom’ decorria de seu caráter valorativo. Pelo contrário, em outros trabalhos ele parece negar o problema do ser e do dever

ser, pois se mostra otimista em relação à possibilidade de reduzir afirmações valorativas a afirmações fáticas:

I think that another perfectly good and natural use of the phrase 'mere statement of fact' is such that it would be contradictory to say of a statement which was 'reducible to factual or existential statements' that it was not a 'mere statement of fact.' And it is by no means clear that statements of the form 'I ought to do X' are not reducible to factual or existential statements, nor, therefore, that they are not, in this sense, 'mere statements of fact.' This, therefore, is a use of 'mere statements of fact' in which 'p is a mere statement of fact' is not identical with 'p is not normative.' It is quite clear that statements of the form 'I ought to do X' are normative, but not clear that they are not 'reducible to factual or existential statements'(Moore, 1968, p.568).

Portanto, a falácia naturalista diz respeito à impossibilidade de definir certas propriedades, independentemente do fato de elas serem valorativas ou não. A impossibilidade de defini-las decorreria do fato de elas não serem compostas por partes, o que inviabilizaria qualquer análise 'decomposicional.' Moore acreditava que cometia um equívoco qualquer um que empregasse sentenças como 'X é bom', a fim de postular a existência de uma identidade entre 'X' e 'bom', embora não houvesse qualquer problema em utilizar sentenças desse tipo quando 'bom' é utilizado apenas como um predicado. A falácia naturalista de Moore consiste justamente nisso, confundir uma predicação com uma definição: não há problemas em dizer algo sobre o 'bom', mas é inaceitável dizer o que é o 'bom', pois ele é indefinível (Joyce, 2006, p. 147-8).

Para sustentar a sua tese, Moore elaborou o argumento da questão em aberto. Esse argumento consiste de um simples teste. Basta adotar qualquer explicação naturalista, como, *e.g.*, alguma que identifique ser bom e ser prazeroso, e, então, elaborar uma questão do seguinte tipo: 'tendo em vista que X é prazeroso, X é bom?' 'Se admitirmos que essa é uma questão em aberto, ou seja, que ela é compreensível e não revela nenhum tipo de confusão conceitual, teremos de admitir que pensar que algo é prazeroso não equivale a pensar que algo é bom: se as duas propriedades fossem idênticas, perguntar se algo que é prazeroso é bom seria tão sem sentido quanto perguntar se algo prazeroso é prazeroso (Sayre-McCord, 2014; Augsberg, 2008, p. 467).

O argumento de Moore teve grande impacto e estimulou o surgimento de explicações não-naturalistas a respeito da moralidade, ou seja, que resistiam à tentação de identificar propriedades morais com propriedades naturais (Sayre-McCord, 2014). Nesse sentido, proliferaram explicações não cognitivistas após a publicação de *Principia Ethica*. De acordo com os filósofos não cognitivistas, os julgamentos morais não seriam expressões descritivas. Segundo eles, a questão sobre se o que é prazeroso é bom permanece em aberto não porque estejamos atribuindo alguma propriedade não natural a certas coisas, mas porque não estamos

atribuindo propriedade alguma. Dizer que algo é bom não equivaleria a fazer uma descrição que pode ser verdadeira ou falsa, nem a manifestar uma crença; equivaleria, na verdade, a manifestar uma atitude conativa (emoções, desejos, prescrições, *etc.*). O discurso moral seria uma ferramenta para influenciar a conduta de outras pessoas, não um instrumento direcionado à busca da verdade (Sayre-McCord, 2014; van Roojen, 2015).

Apesar de ter exercido grande influência, o argumento de Moore apresenta algumas falhas importantes. Em primeiro lugar, precisamos distinguir entre dois possíveis significados das definições naturalistas de ‘bom.’ Quando o naturalista diz, *e.g.*, que ‘para qualquer X, X é bom se e somente se X é prazeroso’, ele pode estar afirmando uma relação linguística *a priori*. Esse é o tipo de relação presente, *e.g.*, na afirmação ‘para qualquer X, X é solteiro se e somente se X for um homem não casado.’ Por outro lado, o naturalista pode estar afirmando uma relação *a posteriori* do tipo ‘para qualquer X, X é água se e somente se X é H₂O’ (Joyce, 2006, p. 149). Ou seja, precisamos distinguir entre o que Michael Smith chama, respectivamente, de *naturalismo definicional* e *naturalismo metafísico* (Smith, 1994, p. 27 e 39).

Quando temos em mente o naturalismo metafísico, o argumento da questão em aberto claramente não funciona. Diante desse tipo de naturalismo, a introspecção não pode servir como um teste de identidade (Joyce, 2006, p. 151). Ao elaborar o seu argumento, Moore não levou em consideração a existência de identidades sintéticas, em oposição às identidades analíticas. Sentenças utilizadas para expressar identidades não precisam ser verdadeiras por definição. Nós fomos capazes, *e.g.*, de descobrir que água é H₂O, no entanto não é possível inferir essa identidade pela simples análise do significado de cada um desses termos (Kripke, 2001). Pensar que um líquido é H₂O é diferente de pensar que ele é água, na medida em que seria perfeitamente inteligível perguntarmos se um líquido que concordamos ser água é de fato H₂O. Basta lembrar que até o Século XVIII não sabíamos que água é H₂O. Ainda assim, *ser* água é idêntico a *ser* H₂O. Logo, a existência de diferenças na forma como pensamos as propriedades não implica a existência de uma real diferença na natureza dessas propriedades. Assim, o argumento de Moore poderia servir para demonstrar que pensar que algo é bom é diferente de pensar que algo é prazeroso, mas não demonstraria que ser bom é diferente de ser prazeroso (Sayre-McCord, 2014; van Roojen, 2015).

Ainda assim, poderíamos ser caridosos e interpretar que o ataque de Moore é direcionado apenas ao naturalismo definicional. No entanto, além de deixar incólume todas as afirmações naturalistas metafísicas, o argumento da questão em aberto ainda enfrentaria sérios problemas (Joyce, 2006, p. 149-151). O principal deles diz respeito ao paradoxo da análise. A prática da análise conceitual pretende oferecer alguma contribuição para o esclarecimento de questões

filosóficas. No entanto, se todas as verdades analíticas precisam ser tão óbvias quanto, *e.g.*, a identidade existente entre a propriedade de ser solteiro e a de ser um homem não casado, a análise conceitual necessariamente resultaria em soluções triviais ou então em soluções falsas — conclusão que seria muito embaraçosa para os filósofos analíticos (Smith, 1994, p. 37-9). Por outro lado, se admitirmos que verdades analíticas não precisam ser óbvias, então o argumento da questão em aberto perde a sua força. Afinal, do fato de não ser óbvio para um falante competente que a propriedade de ser bom é idêntica à propriedade de ser prazeroso não podemos deduzir que uma análise conceitual não nos levaria à conclusão de que elas são propriedades idênticas (Lenman, 2014; van Roojen, 2015). Em suma, uma verdade *a priori* não precisa ser óbvia, ela pode ser demonstrada em certos casos apenas após bastante trabalho intelectual (Joyce, 2006, p. 149).

Além disso, o argumento de Moore envolve ainda uma petição de princípio, na medida em que questões do tipo ‘tendo em vista que X é prazeroso, X é bom?’ só seriam questões em aberto quando a definição ‘bom’ é idêntico a ‘prazeroso’ fosse falsa, e a falsidade desse tipo de identidade é exatamente o que Moore está tentando demonstrar (Frankena, 1939, p. 473).

Se compreendermos o naturalismo moral como a afirmação de que fatos e propriedades morais são fatos e propriedades naturais, as críticas ao argumento da questão em aberto demonstram que ele não serve para refutar *a priori* o projeto dos naturalistas morais. Não está claro por qual razão o naturalista necessitaria oferecer algo como uma definição decomposicional de ‘bom’, e, ainda que ele pretendesse oferecer tal definição, a única objeção que Moore oferece é o argumento da questão em aberto, com todas os problemas vistos (Joyce, 2006, p. 151-2, Lenman, 2014).

Apesar disso, alguns autores, como Terence Horgan e Mark Timmons (Horgan & Timmons, 1992; Timmons, 1999), ainda defendem o argumento da questão em aberto e o utilizam para atacar naturalistas morais contemporâneos. No entanto, não é necessário nos aprofundarmos ainda mais nas minúcias envolvidas nas discussões acerca desse argumento. O que me interessa é apenas demonstrar que a ideia de ‘falácia naturalista’, no sentido utilizado por Moore, é muito mais controversa do que sugere a sua frequente utilização, e, ainda que o argumento da questão em aberto possa ter alguma força contra o naturalismo moral, ele não é relevante contra o tipo de abordagem naturalista adotada nesta pesquisa.

1.2 A falácia naturalista de Hume: a passagem do ‘ser’ para o ‘dever ser’

Embora inicialmente tenha sido um equívoco utilizar a expressão ‘falácia naturalista’ para censurar tentativas de derivar normas de fatos, esse equívoco já foi cometido tantas vezes

que é possível conceder, com base na ideia de que se um número suficiente de pessoas comete um erro linguístico ele deixa de ser um erro, que é esse o atual significado da expressão (Joyce, 2006, p. 152).

Hoje em dia, quando ouvimos essa expressão, é provável que estejamos diante de uma referência ao problema da passagem do ‘ser’ para o ‘dever ser.’ Esse problema foi identificado por Hume neste famoso trecho do *Tratado da Natureza Humana*:

In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprized to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, it is necessary that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it (2009/1740, p. 715)

Há pelo menos duas formas de interpretar essa passagem. De acordo com a primeira delas, comete a ‘falácia naturalista’, ou viola a ‘lei de Hume’, quem tenta deduzir de premissas puramente descritivas uma conclusão normativa. De acordo com a segunda interpretação, muitas vezes adotada, qualquer forma de transição do ‘ser’ para o ‘dever ser’ é ilegítima.

Tratar desse problema aqui é importante por duas razões. Por um lado, há aqueles que imputam a qualquer um que procure descrever comportamentos humanos a acusação de estar de alguma forma justificando-os. Por trás desse tipo de imputação, costuma estar presente o que Bischof (2012, p. 16) denomina ‘falácia moralista’, *i.e.*, a dedução de uma conclusão fática a partir de premissas normativas: Se A não deve ser verdade, então A não é verdade. É comum, por exemplo, dentre aqueles que defendem a atribuição de direitos iguais para homens e mulheres a prática de negar de antemão qualquer estudo na área de Biologia que aponte diferenças comportamentais entre homens e mulheres. Esse tipo de erro é uma simples inversão do erro cometido por aqueles que, a partir de referências às diferenças biológicas entre homens e mulheres, deduzem imediatamente que deve haver alguma diferença na atribuição de direitos. A conclusão do ‘ser’ a partir do ‘dever ser’ é tão ilegítima quanto a conclusão do ‘dever ser’ a partir do ‘ser’ (Bischof, 2012, p.16).

Por outro lado, qualquer um que procure compreender a moralidade a partir de ciências biológicas está sujeito à acusação de praticar o erro grosseiro de derivar valores de fatos. De acordo com a acusação, a moralidade nos diz o que devemos fazer, enquanto a Biologia e outras ciências apenas descrevem como as coisas são. Deste modo, pesquisas como esta seriam inúteis para discussões a respeito da moralidade, caso se restringissem a meras descrições, ou envolveriam uma falácia, caso pretendessem obter alguma conclusão valorativa.

1.2.1 Primeira interpretação da lei de Hume

Conforme a primeira interpretação mencionada, afirmações sobre fatos são formalmente distintas de afirmações valorativas, de modo que não é possível deduzir de modo válido uma norma ou qualquer conclusão valorativa somente a partir premissas fáticas, *i.e.*, nenhum conjunto de premissas no qual estejam presentes apenas premissas puramente descritivas é suficiente para chegarmos a uma conclusão valorativa (van Roojen, 2015; Sayre-McCord, 2014; Searle, 1964, p. 43).

A tese da impossibilidade de deduzir normas de descrições fáticas, de acordo com essa interpretação, não implica que entre as premissas que levam a uma conclusão normativa não possam estar presentes também premissas de natureza descritiva. O seguinte argumento, *e.g.*, que apresenta uma premissa descritiva e uma conclusão normativa, é válido: quem não trabalha, não deve comer; A não trabalha; Logo, A não deve comer (Hoerster, 1969, p. 14-15).

O que a tese implica é que somente é possível deduzir uma conclusão normativa se, além das premissas descritivas, houver pelo menos uma premissa valorativa. Portanto, se Hume estiver certo, qualquer argumento válido que contenha uma conclusão normativa inclui ou pressupõe alguma premissa valorativa e o erro estaria em não a explicitar. Quem, por exemplo, deduz do fato de que alguém está sentindo dor que algo de ruim está acontecendo, está pressupondo que a dor é algo ruim (Hoerster, 1969, p. 14-15; Sayre-McCord, 2014).

O problema dessa advertência de Hume é que ela tem escopo bastante limitado. Como salienta Birnbacher (2003, p. 364, *apud* Augsberg, 2008, p. 472), o argumento de Hume só é aplicável quando duas condições bastante restritivas estão presentes: (a) a tentativa de derivar uma norma (dever ser) de um fato (ser) pretende ser estritamente dedutiva; (b) as premissas a partir das quais a norma é derivada são compreendidas em um sentido puramente descritivo. No entanto, grande parte das tentativas de derivar normas de fatos não preenchem essas duas condições, de forma que elas não são, a rigor, falaciosas nesse sentido.

Há várias maneiras de contornar o problema da transição do ser para o dever ser. Basta incluir de algum modo um elemento valorativo entre as premissas e a objeção de Hume perde a sua força. Essa inclusão muitas vezes ocorre de forma pouco perceptível, em razão da separação entre conceitos normativos e descritivos não ser tão aguda. Há uma grande zona cinzenta na qual estão presentes conceitos que misturam elementos descritivos e valorativos. Quando se derivam conclusões claramente normativas a partir de premissas nas quais estão presentes conceitos que apresentam de forma velada elementos valorativos, a lei de Hume não se aplica (Augsberg, 2008, p. 470-2).

Em outros casos, os elementos valorativos presentes entre as premissas decorrem da observância dos objetivos e desejos do agente. Mackie menciona diversos exemplos nos quais não há problemas na transição do ‘ser’ para o ‘dever ser.’ Um deles é o jogo de xadrez. Quando consideramos as regras do jogo e a posição atual das peças em conjunto com o objetivo de vencer o jogo, podemos concluir quais movimentos o jogador pode fazer e quais ele *deve* fazer. Mackie salienta como em diversos outros contextos não é difícil chegar à conclusão de um ‘dever ser’, entendido aqui como um imperativo hipotético, a partir de premissas fáticas:

If someone wants to get to London by twelve o'clock, and the only available means of transport that will get him there is the ten-twenty train, and catching this train will not conflict with any equally strong desires or purposes that he has then he ought to indeed must, catch the ten-twenty (1990, p. 65-66).

Quando se descreve de forma satisfatória as condições fáticas acerca do agente, incluindo os seus desejos, chega-se à conclusão sobre qual ação ele deve tomar. Dizer que o agente deve agir de tal forma, significa dizer que ele tem alguma razão para agir dessa forma. Os objetivos do agente somados às relações causais conhecidas constituem a razão para ele agir (Mackie, 1990, p. 65-66).

Por outro lado, há autores, como John Searle, que acreditam que o problema não precisa ser sequer contornado, ou seja, acreditam na possibilidade de derivar uma conclusão valorativa a partir de uma descrição puramente fática. Ele propõe o seguinte exemplo:

- (1) Jones uttered the words “I hereby promise to pay you, Smith, five dollars.”
- (2) Jones promised to pay Smith five dollars.
- (3) Jones placed himself under (undertook) an obligation to pay Smith five dollars.
- (4) Jones is under an obligation to pay Smith five dollars.
- (5) Jones ought to pay Smith five dollars (Searle, 1964, p. 44).

Searle sugere que, nesse exemplo, a lacuna entre o ‘ser’ e o ‘dever ser’ pode ser preenchida por um tipo específico de fato, os chamados ‘fatos institucionais.’ Ele considera que a palavra ‘promessa’ possui um caráter valorativo na medida em que a noção de promessa está vinculada à noção de obrigação. No entanto, Searle também considera que essa noção é puramente descritiva, pois é uma questão objetiva se alguém fez ou não uma promessa. Por essa razão, segundo Searle, a distinção entre valorativo e descritivo deveria ser reexaminada (Searle, 1964, 57-8; Mackie, 1990, p. 67-72).

De acordo com Searle, a aceitação da distinção rígida entre descrição e valoração tem como origem uma visão equivocada acerca de como as palavras se relacionam com o mundo. Conforme essa visão, seriam exemplos de afirmações descritivas sentenças como: “meu carro se move a 80 milhas por hora”; “Jones mede 1,83m de altura”; “Smith tem cabelos castanhos.” Exemplos de afirmações valorativas seriam: “meu carro é bom”; “Jones deve pagar 5 dólares a Smith”; “Smith é um homem repugnante.” A veracidade ou a falsidade das afirmações do

primeiro tipo seria verificável de modo objetivo, pois o conhecimento do significado das expressões descritivas implicaria o conhecimento das circunstâncias objetivamente verificáveis sob as quais as afirmações por elas compostas são verdadeiras ou falsas. No entanto, em relação às afirmações valorativas a situação seria diferente. Qualquer justificação de afirmações valorativas acabaria envolvendo o apelo às atitudes, aos critérios de avaliação ou aos princípios morais adotados por quem as pronuncia. Desta forma, afirmações descritivas seriam objetivas, enquanto afirmações valorativas seriam subjetivas. Essa diferença existiria em razão de afirmações valorativas cumprirem uma função distinta daquela desempenhada pelas afirmações descritivas. A função daquelas seria expressar emoções e atitudes, condenar, elogiar, recomendar, advertir, *etc.* (visão não cognitivista). A partir da constatação dessa diferença entre afirmações valorativas e descritivas, ficaria clara a existência de uma espécie de abismo lógico separando-as (Searle, 1964, 52-3).

Segundo Searle, um dos problemas a respeito dessa visão é a sua incapacidade de distinguir entre dois tipos de afirmações descritivas: as sobre fatos brutos e as sobre fatos institucionais. Questões sobre (1) se alguém se casou, se fez uma promessa, se possui cinco dólares são exemplos de questões sobre fatos objetivos, assim como o são as questões sobre (2) se alguém tem cabelos ruivos ou olhos castanhos. No entanto, há uma diferença importante entre esses dois tipos de questões. As questões do primeiro tipo envolvem expressões como ‘casamento’, ‘promessa’ e ‘cinco dólares.’ Expressões como essas fazem referências a fatos cuja existência pressupõe certas instituições: “um homem tem cinco dólares, dada a existência da instituição do dinheiro. Elimine essa instituição e tudo que ele tem é um pedaço retangular de papel com tinta verde.” De modo semelhante, só podemos nos casar ou fazer promessas em um contexto no qual as instituições do casamento e da promessa existem. Sem essas instituições, estaríamos apenas proferindo palavras e gesticulando. Os fatos desse primeiro tipo poderiam ser caracterizados como fatos institucionais em contraste com os fatos do segundo tipo: “um homem possuir um pedaço de papel com tinta verde é um fato bruto, ele ter cinco dólares é um fato institucional” (Searle, 1964, 52-6, tradução livre).

Quando Searle emprega a palavra instituição, ele tem em mente um sistema de regras constitutivas. Searle faz uma distinção entre regras regulativas e regras constitutivas. Regras do primeiro tipo regulam formas de comportamento já existentes antes da criação da regra. Algumas regras de etiqueta, *e.g.*, disciplinam o modo como comemos, mas a atividade de comer existe independentemente da existência dessas regras. Outras regras, entretanto, não apenas regulam, mas criam novas formas de comportamento. As regras de xadrez, *e.g.*, criam a

atividade de jogar xadrez e a definem. O jogo de xadrez não existe sem a existência dessas regras (Searle, 1964, 52-6).

As instituições do casamento, do dinheiro e da promessa, assim como o xadrez, seriam formadas por regras desse último tipo. Conforme Searle, uma vez que tenhamos reconhecido a existência desses fatos institucionais, *i.e.*, de fatos cuja existência depende de um sistema de regras constitutivas, não seria difícil reconhecer que muitas formas de obrigações, de direitos e de responsabilidades são institucionalizadas de modo semelhante. Dessa forma, em muitos casos seria uma questão de fato a existência de certas obrigações, direitos e responsabilidades, “mas uma questão de fatos institucionais, não de fatos brutos” (Searle, 1964, 52-6, tradução livre). Portanto, dentro de um sistema de regras constitutivas seria sim possível derivar um ‘dever ser’ a partir de fatos.

No entanto, Mackie destaca como a explicação de Searle se vale de uma ambiguidade a respeito de como falamos sobre instituições. Nós podemos descrevê-las adotando uma perspectiva externa. Nesse sentido uma instituição poderia ser assim descrita: há uma instituição ou prática social chamada promessa. Parte dessa instituição envolve certas demandas daqueles que dela participam. No caso da instituição da promessa, existe uma demanda para que as promessas sejam mantidas – na instituição do xadrez, de forma análoga, há uma demanda para que apenas certos movimentos sejam feitos. Por outro lado, podemos falar sobre a instituição da promessa como se dela fizessemos parte, *i.e.*, podemos adotar uma perspectiva interna. Logo, poderíamos dizer simplesmente ‘Jones deve pagar cinco dólares a Smith’ (porque ele prometeu agir dessa maneira), assim como alguém que estivesse jogando xadrez poderia falar ‘você não pode mover aquela torre’ (porque isso deixaria o seu rei em xeque) (Mackie, 1990, p. 67).

Mackie então questiona em qual dos dois sentidos devemos interpretar a conclusão do argumento de Searle (*Jones deve pagar cinco dólares a Smith*). Se a interpretarmos a partir de uma perspectiva externa, ela quer dizer algo como ‘a instituição da promessa demanda que Jones pague cinco dólares a Smith.’ Caso adotemos essa interpretação, qualquer um que tenha aceito a primeira premissa do argumento (*Jones proferiu as palavras ‘eu por meio deste ato prometo pagar a você, Smith, cinco dólares’*) e a descrição das circunstâncias envolvidas terá de aceitar a conclusão proposta por Searle. Contudo, quando interpretada dessa forma, a conclusão deixa de ser valorativa ou prescritiva e, portanto, deixa de contrariar a lei de Hume (Mackie, 1990, p. 67-8).

Por outro lado, se interpretarmos a conclusão a partir de uma perspectiva interna, teremos de admitir que há uma importante transição na passagem de (2) — *Jones prometeu pagar cinco dólares a Smith* — para (3) — *Jones contraiu uma obrigação de pagar cinco*

dólares. O que se segue de (2), quando a interpretamos a partir de uma perspectiva externa, é que Jones manifestou a intenção de contrair uma obrigação. No entanto, apenas quando invocamos as regras da instituição da promessa (em contraposição à atitude de somente reportá-las) podemos inferir que Jones de fato contraiu uma obrigação. Nesse caso, o argumento é válido não pelas regras gerais da lógica, mas por uma lógica especial, utilizada quando raciocinamos ‘dentro’ da instituição da promessa (Mackie, 1990, p. 67-8).

De acordo com Mackie, certas palavras, como ‘promessa’, quando são utilizadas no âmbito de suas respectivas instituições, apresentam uma característica lógica peculiar:

The performance of a certain speech act in appropriate circumstances is, in virtue of one part of the meaning of the words, sufficient to validate the statement ‘Jones promised ...’ (...). But also, in virtue of another part of their meaning, such a statement entails ‘Jones ought to ...’ or ‘John has a right to ...’ The very meanings of these words thus embody synthetic claims, in fact justifications of transitions from ‘is’ to ‘ought.’ For this reason the adopting of such fragments of language is not a neutral matter, as is the use of most parts of language, such as figure in most arguments: to use the word ‘promise’ (...) with its full within-the-institution meaning is already to endorse the institution in a substantial way, to adopt and support certain distinctive patterns of behaviour and to condemn others (Mackie, 1990, p. 71-2).

Portanto, quando interpretamos a conclusão dessa segunda maneira, adotando a perspectiva interna, ela é valorativa e prescritiva, mas só pode ser derivada de (1) e das demais premissas fáticas quando empregamos um tipo especial de lógica, ou seja, somente quando adotamos as regras da instituição da promessa. Conforme Mackie, ninguém que esteja preocupado com o ‘espírito’ da lei de Hume, em vez de sua literalidade, precisa se preocupar com uma derivação desse tipo (1990, p. 67-8). O argumento de Searle serviria apenas para demonstrar que dentro de um contexto específico, quando pensamos no interior de uma instituição, podemos sim derivar normas de fatos.

Em suma, quando interpretada dessa primeira forma – como a afirmação da impossibilidade de deduzir uma conclusão normativa de premissas puramente descritivas — e tendo em vista as ressalvas feitas quanto às possibilidades de contornar o problema por meio da inclusão dos objetivos do agente nas premissas ou por meio do apelo a certas instituições, a lei de Hume é bastante plausível.

1.2.2 Segunda interpretação da Lei de Hume

Conforme a segunda interpretação, qualquer forma de transição do ‘ser’ para o ‘dever ser’ é ilegítima. Nenhuma norma poderia ser fundamentada com base em fatos biológicos, e qualquer tentativa de misturar elementos descritivos e normativos em um argumento lógico dedutivo seria rotulada como uma falácia. Logo, estudos científicos seriam irrelevantes para as discussões éticas. Ainda que esses estudos possam nos ajudar a identificar quais são os

princípios morais que as pessoas adotam em seus julgamentos ou quais os princípios que elas seriam capazes de aceitar, eles não poderiam dar qualquer contribuição para a tarefa de encontrar os princípios corretos. Afinal, se a ciência apenas descreve fatos, ela não poderia nos dizer nada acerca do que é certo ou errado (Augsberg, 2008, p. 470-2; Ritchie, 2008, p. 6-7).

Há diversos autores que adotam essa posição e consideram os dados empíricos irrelevantes para a formulação de teorias normativas. Para eles haveria uma independência entre o ‘ser’ e o ‘dever ser’, de modo que tentativas de tratar da moralidade por meio de uma abordagem empírica incorreriam no erro de misturar dois domínios incompatíveis (Bischof, 2012, p. 14). Portanto, a pesquisa empreendida nesta dissertação seria irrelevante para discussões normativas, pois, não importa o quanto saibamos sobre como as coisas de fato são, isso não nos diria nada a respeito de como elas devem ser.

Embora essa segunda interpretação tenha um escopo muito mais amplo, ela é também menos plausível. Uma das razões pelas quais essa interpretação não parece correta é o fato de que Hume, em sua própria teoria ética, deriva o ‘dever ser’ a partir do ‘ser’, o que indica que ele não considera essa transição proibida. Hume acreditava que o comportamento moral se apoia na motivação social, no nosso desejo de conviver em grupo e de ser admirado. Essa motivação seria parte de nossa natureza biológica. Hume buscou fundamentar a moralidade a partir daquilo que somos e das coisas com as quais nos preocupamos. A intenção dele não era negar a relevância de considerações fáticas nos nossos juízos acerca do que devemos fazer, mas sim distanciar-se de inferências esdrúxulas do tipo ‘os maridos são mais fortes do que suas esposas, logo as esposas devem obedecer aos seus maridos’ (Churchland, 2011, p. 4-7).

De qualquer forma, ainda que essa segunda interpretação da lei de Hume não corresponda ao que o filósofo quis dizer, alguém poderia sustentar que a tese derivada dessa interpretação está correta, ou seja, que os domínios do ‘ser’ e do ‘dever ser’ são de fato incomunicáveis. Contudo, essa tese é também bastante implausível.

Em primeiro lugar, teorias normativas, a fim de que elas possam ser proveitosas, dependem de observações empíricas. Não há sentido em recomendar condutas que as pessoas são incapazes de executar, ou seja, ‘dever’ implica ‘poder’ (Abrantes, 1994, p. 184; Sober, 2000, p. 199). A menos que as teorias éticas sejam formuladas com o objetivo de fornecer conselhos inúteis, elas deveriam levar em consideração os seres humanos reais aos quais elas se destinam, e não seres ideais, com capacidades físicas e psicológicas ilimitadas. Se o princípio de que ‘dever’ implica ‘poder’ estiver correto, resultados de estudos científicos podem, eventualmente, demonstrar que várias afirmações a respeito de como as pessoas devem agir são descabidas (Sober, 2000, p. 199). Estudos empíricos sobre moralidade servem para fornecer uma espécie de

‘fotografia’ dos valores vigentes nos grupos humanos, uma análise a respeito da origem desses valores e, conseqüentemente, prognósticos a respeito das possibilidades e dos métodos para alterá-los. Qualquer teoria normativa que busque alguma aplicação prática deve levar em conta esses fatores.

Apesar disso, há não muito tempo, a grande maioria dos filósofos simplesmente ignorava a importância das descobertas científicas para as discussões sobre julgamentos morais. Nas últimas décadas, no entanto, cada vez mais filósofos procuram aprender sobre as ciências dedicadas ao estudo da mente e do cérebro e apresentam os resultados de experimentos conduzidos por cientistas para sustentar as suas posições filosóficas (Hauser, 2007, p. 440). Esses filósofos contemporâneos percebem que há uma sobreposição entre as áreas de interesse da Ciência e da Filosofia, e nessas áreas onde elas se cruzam a Ciência tem produzido resultados muito mais proveitosos do que a Filosofia. Entretanto, há áreas nas quais a Ciência não pode oferecer as respostas. Ela não pode nos dizer, *e.g.*, como devemos viver as nossas vidas ou o que devemos valorizar. Mas, mesmo nessas áreas, a Ciência impõe certas restrições, na medida em que qualquer teoria que pretenda ser respeitada não deve contradizer o conhecimento científico estabelecido (Ritchie, 2008, p. 200). O tipo de naturalismo que adoto aqui implica que as teorias científicas aceitas devem funcionar como uma espécie de restrição sobre quais teorias éticas e políticas podemos elaborar; ou seja, parto da ideia de que deve haver consistência (consiliência) entre as teorias éticas e científicas (Ritchie, 2008, p. 107).

Além disso, embora implicações normativas não possam ser extraídas apenas de alegações científicas, quando combinamos essas alegações com outras premissas normativas triviais podemos chegar a conclusões normativas importantes. Por exemplo, descobriu-se que decisões sobre a aplicação de pena de morte proferidas por júris são algumas vezes afetadas por vieses racistas. Se combinarmos essa descoberta com a alegação normativa de que júris não devem ser afetados por esse tipo de viés, concluiremos que algumas decisões deveriam ter sido tomadas de outro modo (Greene, 2014, p. 711).

No nosso dia a dia, frequentemente pensamos dessa maneira, combinando juízos normativos com juízos descritivos, e não há qualquer falha lógica nesse tipo de raciocínio. Cada vez que enfrentamos algum problema, empregamos a nossa capacidade de percepção e de compreensão, o nosso conhecimento e as nossas emoções para inferir como devemos agir. Frequentemente decidimos o que fazer com base nos fatos envolvidos no caso, nas nossas experiências passadas e nos nossos desejos. Se, por exemplo, tenho uma dor de dente muito forte, concluo que devo ir ao dentista (Churchland, 2011, p. 4-7).

Em suma, rejeito a tese decorrente da segunda interpretação da lei de Hume, pois questões fáticas são relevantes para discussões normativas, ainda que não seja possível derivar dedutivamente conclusões normativas apenas de premissas puramente fáticas. Com relação à primeira interpretação para essa lei, ela pode ser aceita, mas com a ressalva de que tem um escopo de aplicação muito limitado e que não serve como prova irrefutável contra o naturalismo moral, na medida em que a violação da lei de Hume não implica nenhuma falácia.

1.2.3 Relevância da advertência de Hume

Uma das possíveis lições que podemos tirar do argumento de Hume é que sempre que premissas extraídas da biologia, ou conforme a ‘natureza’, são utilizadas para defender conclusões éticas, há ao menos uma premissa valorativa, ainda que implícita⁶. Quando essa premissa pressuposta é explicitada, os argumentos muitas vezes se mostram implausíveis (Sober, 2000, p. 209). A teologia moral católica, *e.g.*, proíbe o sexo com o uso de preservativos, pois, de acordo com a ‘natureza’, o ato sexual serviria apenas para a procriação. Nesse argumento, está subentendida a premissa de que ‘não devemos contrariar a natureza’ (Bischof, 2012, p. 15). Embora essa premissa tenha algum apelo popular, ela é simplória e insustentável: da constatação que a violência é algo natural, não decorre que devemos ser violentos; da constatação que a xenofobia é natural entre os seres humanos não decorre que a hostilidade a estrangeiros deva ser tolerada, *etc.* (Joyce, 2006, p. 152).

Contudo, o fato de a premissa ser falsa não torna o argumento falacioso. Uma falácia é uma falha no raciocínio manifestada na forma do argumento, não uma falsa crença. Quando alguém diz que a homossexualidade não é natural, que todas as coisas não naturais são erradas e que, portanto, a homossexualidade é errada, não há qualquer falácia, pois o raciocínio é perfeito (Joyce, 2006, p. 152). O problema desse argumento não está na sua estrutura lógica, mas nas suas premissas, tendo em vista que há inúmeros traços presentes na ‘natureza humana’ que são bastante questionáveis sob o prisma moral.

Justamente em razão de esse tipo de argumento não apresentar falhas na sua estrutura lógica, a lei de Hume não serve como prova irrefutável contra o naturalismo moral. Nesse sentido, Joyce apresenta o seguinte exemplo. Suponhamos que algum naturalista moral sustente que alguma propriedade P surgida ao longo da evolução humana equivale à propriedade de ser moralmente correto (‘moral oughtness’) e proponha o seguinte argumento:

⁶ Essa é outra das razões pelas quais a expressão falácia naturalista não é adequada para fazer referência ao problema da passagem do ser para o dever ser, pois, em geral, trata-se apenas de um entimema, não de uma real falácia (Bischof, 2012, p. 15).

*P1 - A ação X de Pedro tem a propriedade P;
Logo, Pedro deve praticar a ação X.*

Quando formulado dessa forma o argumento é inválido. No entanto, esse argumento pode ser tratado como um entimema no qual está pressuposta a seguinte premissa:

P2- Para qualquer ação X, X tem P se e somente se X deve ser praticada.

Certamente P2 será controversa, mas essa não é a questão. O naturalista moral acredita que ela é verdadeira, não que ela é incontestável. Afirmar que P2 não pode ser verdadeira simplesmente porque conecta uma afirmação descritiva (ser) a outra normativa (dever ser) é uma petição de princípio. O naturalista moral terá trabalho para nos convencer de que P2 é verdadeira, mas o seu projeto não envolve necessariamente alguma falácia, nem está *a priori* fadado ao fracasso (Joyce, 2006, p. 155).

Se o naturalista moral estiver correto, o problema da passagem do ‘ser’ para o ‘dever ser’ desaparece. O mesmo raciocínio se aplica para outras formas de realismo comprometidas com a tese de que propriedades morais estão presentes no mundo para serem descobertas, independentemente das preocupações e práticas dos seres humanos. Se estas posições estiverem erradas, o que me parece provável, teríamos de conceber as normas morais como imperativos hipotéticos ou então como imperativos não-hipotéticos derivados de certas instituições humanas. Quando concebemos as normas morais dessa forma, os estudos empíricos passam a ter grande valor, pois eles indicam quais são os melhores caminhos para atingirmos os nossos objetivos. Portanto, o apelo à expressão ‘falácia naturalista’ não é suficiente para refutar de antemão qualquer tentativa de elaborar uma teoria ética baseada em premissas fáticas (Augsberg, 2008).

De qualquer modo, a advertência de Hume é útil em razão de chamar atenção para o fato de que a transição do ‘ser’ para o ‘dever ser’ é problemática e exige cuidado na hora de esclarecê-la e justificá-la (Hoerster, 1969, p. 12-13). Será relevante para a interpretação dos próximos capítulos manter em mente essa advertência.

No quarto capítulo, por exemplo, examinaremos a teoria da Gramática Moral Universal. Se essa teoria estiver correta, e os nossos julgamentos morais forem de fato produzidos a partir dos princípios universais dessa gramática, algum teórico poderia sugerir que deveríamos aplicar de modo consciente esses princípios, evitando fazer julgamentos em desacordo com eles. Em um caso como esse, poderíamos questionar qual a fundamentação que esse teórico emprega para chegar a essa conclusão normativa. Caso ele justificasse a afirmação de que ‘devemos aplicar de forma consciente os princípios da Gramática Universal’ por meio de um apelo aos princípios da Gramática Universal, ele claramente estaria argumentando de forma circular. Para extrairmos implicações normativas desse suposto fato, seria necessário explicar por quais motivos

deveríamos seguir os princípios dessa gramática. Por que seria mais legítimo segui-los do que tentar bloqueá-los? Quais os motivos que teríamos para achar que essa gramática, supostamente surgida em razão de pressões seletivas existentes em ambientes bastante diferentes daqueles nos quais vivemos hoje, não possui princípios que, quando aplicados nas sociedades contemporâneas, resultam na produção de conflitos, de sofrimento e de outras coisas que preferiríamos evitar (Bruni, Mameli & Rini, 2014, p. 162)? Conhecer a Gramática Moral Universal não implicaria diretamente qualquer ‘dever ser’, nem mesmo em um sentido prudencial.

Com relação às demais teorias que veremos a seguir acerca da origem e do desenvolvimento da moralidade é preciso ter esse mesmo cuidado. Na presente pesquisa busco examinar os argumentos e os estudos científicos utilizados para sustentar a veracidade ou a falsidade dessas teorias. Caso alguma delas seja verdadeira, ou seja considerada a melhor explicação disponível sobre como fazemos os nossos julgamentos morais, disso não se segue diretamente qualquer conclusão normativa.

2 Nativismo moral: aspectos conceituais

Há uma tradição na Filosofia, que remonta pelo menos até Platão, de se considerar que uma definição correta deve especificar as condições que são necessárias, quando consideradas de forma isolada, e suficientes, quando consideradas em conjunto, para a utilização do conceito a ser definido. Assim, uma definição deveria especificar o conjunto de características presente em todos os objetos aos quais o conceito é aplicável e apenas neles. Se for possível encontrar qualquer exceção em relação à definição, ela está errada. Ou seja, se houver objetos que se encaixam na definição, mas aos quais o conceito não se aplica, ou objetos que não se encaixam na definição, mas aos quais o conceito se aplica, então a definição deve ser revista (Stich, 2012, p. 88).

No entanto, desde a década de 1970 há uma série de pesquisas que desafiam essa visão clássica acerca dos conceitos (Rips et al., 1973; Rosch & Mervis, 1975; Rosch 1999; Malt & Smith, 1984; Medin & Smith, 1984). Um dos principais problemas com ela diz respeito aos chamados *efeitos de tipicidade*. Essa expressão é utilizada para fazer referência ao fato de que as pessoas são capazes de classificar de modo consistente objetos que se enquadram em determinado conceito quando questionadas a respeito de quão ‘típicos’ ou ‘representativos’ esses objetos são do referido conceito. Em experiências desse tipo, as pessoas categorizam de forma mais rápida e precisa objetos considerados típicos e os mencionam com uma frequência maior quando são instigadas a fornecer exemplos de um determinado conceito, ou seja, elas não atribuem o mesmo *status* a todos os objetos de uma categoria. Assim, por exemplo, em relação ao conceito de pássaro, o canário é considerado um exemplar mais típico do que uma galinha; em relação ao conceito de fruta, um limão é considerado mais típico do que um abacate.

Os resultados dessas experiências indicam que não costumamos categorizar objetos conforme a visão clássica. Se um conceito especifica condições necessárias e suficientes, todos os objetos que satisfazem essas condições subsumem-se igualmente ao conceito em questão. Dessa forma, é difícil de explicar por que alguns objetos são considerados mais típicos, mais fáceis de serem categorizados ou de serem lembrados. Se o conceito de pássaro fosse definido em termos de condições necessárias e suficientes, poderíamos presumir que canários e galinhas preencheriam essas condições igualmente bem. Tendo em vista que não temos contato mais frequente com canários do que temos com galinhas, seria difícil de explicar, a partir da visão clássica, por que motivo os categorizamos mais rapidamente como pássaros (Harman, Mason & Sinnott-Armstrong, 2010, p. 228-30; Stich, 2012, p. 92-4).

Com base nesses estudos sobre os efeitos de typicalidade, foram desenvolvidas novas teorias sobre conceitos contrastantes com a teoria clássica. Dentre elas destacam-se as teorias prototípicas de conceitos e as teorias de conceitos como exemplares (Harman, Mason & Sinnott-Armstrong, 2010, p. 228-30).

As teorias prototípicas de conceitos partem do argumento utilizado por Wittgenstein em *Investigações Filosóficas* para sustentar que os referentes de uma palavra não precisam apresentar uma característica em comum para que a palavra possa ser compreendida e utilizada pelos falantes de determinada língua. Wittgenstein utiliza como exemplo o conceito de jogo. Segundo ele, não há nenhuma característica que esteja presente em todos os diversos tipos de jogos — de tabuleiro, de cartas, de bola, olímpicos, etc. Há apenas uma complexa rede de semelhanças, as quais ele denomina *semelhanças de família* (*Familienähnlichkeiten*), que se sobrepõem e se cruzam em alguma medida. O que faz com que determinada atividade seja considerada um jogo é o fato de ela ter uma quantidade suficiente de características que estão presentes em um número suficiente de outras atividades que se enquadram no conceito. As semelhanças que unem os diversos tipos de jogos seriam como as semelhanças presentes nos membros de uma família: é possível reconhecer os membros de uma família por características como a cor dos olhos e dos cabelos, o temperamento, as feições do rosto, a compleição física, mas nenhuma dessas características está necessariamente presente em todos os membros da família (1986, p. 31-2).

Se alguém nos perguntar quais são os limites exatos do conceito de jogo não saberemos responder. O que poderemos fazer é dar exemplos de jogos, descrever as suas características e acrescentar que costumamos chamar de jogos coisas semelhantes à descrição feita. Em casos como esse, a impossibilidade de oferecermos os limites exatos não decorreria de nossa ignorância, mas do fato de eles não existirem.

Conceitos como o conceito de jogo podem ser pensados como uma figura circunscrita por linhas bastante borradas, de modo que muitas vezes é difícil dizer se um ponto qualquer está do lado de dentro ou do lado de fora da figura. Nós poderíamos, no entanto, com alguma finalidade específica, criar limites precisos para a utilização desses conceitos. Com relação ao conceito de jogo, poderíamos traça-los de diversas formas, mas em todas elas algumas das atividades que costumamos chamar de jogo ficariam de fora, ou, pelo contrário, incluiríamos algumas atividades que não costumamos enquadrar nesse conceito. Ou seja, poderíamos elaborar uma maneira de utilizar o conceito com maior precisão, mas teríamos de pagar um certo preço. Não é, portanto, óbvio que a definição precisa nos atenderá melhor do que aquela um pouco mais vaga. Wittgenstein esclarece que embora seja possível traçar limites rígidos para

a utilização de um conceito, esse tipo de delimitação não é necessária para que ele possa ser utilizado e em muitos casos não é sequer desejável. Ele salienta que muitas vezes, principalmente no âmbito da ética e da estética, quando tentamos desenhar uma figura com limites precisos a partir de uma figura borrada corremos o risco de chegar a um resultado que não guarde qualquer semelhança com a figura original. O grau da semelhança entre essas duas figuras dependerá do quanto são vagas as linhas que circunscrevem a figura original. Podemos imaginar uma figura vermelha borrada com forma retangular. A partir dela é possível desenhar diversos retângulos vermelhos diferentes, mas que sejam parecidos com a figura original. No entanto, em um caso extremo, no qual as cores da figura original se misturem com o plano de fundo de modo a não deixar pistas sobre os limites que as separam, qualquer tentativa de traçar limites precisos será fracassada. Em um caso como esse, poderíamos desenhar um retângulo, um círculo ou até mesmo um coração com base na figura original. Em certo sentido qualquer desses desenhos estaria correto, mas, por outro lado, nenhum deles se assemelharia à figura original.

Posteriormente, estudos acerca dos efeitos de tipicidade trouxeram suporte empírico ao argumento de Wittgenstein. Em alguns desses estudos, como, por exemplo, naquele conduzido por Rosch & Mervis (1975), solicitou-se aos entrevistados que especificassem as propriedades de certos exemplares de uma categoria. Em relação à categoria *veículo*, e.g., estavam listados exemplares como carro, caminhão, ônibus, trem, bicicleta, barco, etc. Os resultados desses estudos confirmaram a hipótese de que quanto mais atributos um exemplar tem em comum com outros membros da mesma categoria, mais ele era considerado um exemplar representativo ou típico dessa categoria. A grande maioria das propriedades mencionadas pelos entrevistados não estava presente em todos os membros, ou seja, eram propriedades consideradas não necessárias para o pertencimento à categoria analisada. No entanto, apesar de não serem consideradas necessárias, o fato de essas propriedades estarem correlacionadas com o tempo que os entrevistados levam para categorizar um exemplar e com o quão típico eles o consideram da categoria sugere que nós utilizamos essas propriedades quando decidimos quais exemplares fazem parte de uma categoria (Medin & Smith, 1984, p. 116).

A concepção prototípica defende que um conceito é uma representação formada pelo conjunto das propriedades normalmente compartilhadas pelos indivíduos aos quais o conceito se aplica. De acordo com essa visão, um indivíduo se enquadra em um conceito quando compartilha uma quantidade suficiente de características com esse protótipo, não sendo necessário que alguma característica específica esteja presente. Quanto maior a similaridade do indivíduo com o protótipo, mais rapidamente ele será reconhecido como um exemplar da categoria (Harman, Mason & Sinnott-Armstrong, 2010, p. 230; Stich, 2012, p. 92-4).

Os conceitos que são descritos pela concepção prototípica também são chamados de “cluster concepts” (Gasking, 1960), na medida em que eles são elaborados a partir de um conjunto (*cluster*) de características comuns, mas que não precisam necessariamente estar todas presentes em uma dada instanciação do conceito (Harman, Mason & Sinnott-Armstrong, 2010, p. 230).

As teorias de conceitos como exemplares (*e.g.*, Medin & Schaffer, 1978), por sua vez, sustentam que a categorização ocorre por meio da comparação com exemplares concretos de determinado conceito. Assim como ocorre com as teorias prototípicas, as teorias do exemplar também afirmam que a categorização ocorre por meio da avaliação da similaridade e elas também não se baseiam na existência de um núcleo de condições suficientes e necessárias. No entanto, conforme as teorias do exemplar, a categorização não é feita pela comparação com protótipos abstratos, mas pela comparação das semelhanças com um exemplar específico ou com um conjunto de exemplares específicos. Assim, por exemplo, eu julgaria se algo é um sofá comparando-o com o sofá que tenho em minha sala de estar ou com outros que já vi ao longo de minha vida. A categorização ocorreria por meio da ativação mental da representação de um ou mais exemplares de um conceito e da posterior avaliação da similaridade entre o objeto a ser categorizado e essa representação. Embora os dois tipos de modelos sejam parecidos, quando desenvolvidos em detalhe, eles produzem previsões diferentes⁷ (Stich, 2012, p. 92-4; Harman, Mason & Sinnott-Armstrong, 2010, p. 231).

Há estudos que indicam que nenhuma dessas teorias é capaz de explicar de forma geral como todos os tipos de conceito são formados. Em alguns casos, as teorias de conceitos como exemplares fazem previsões melhores; em outros, as teorias prototípicas são superiores (Barsalou, 1987). Há ainda certos casos nos quais os entrevistados empregam também critérios teóricos em suas categorizações, de modo que elas não podem ser explicadas exclusivamente por modelos baseados em avaliação de similaridades (Smith, 1995, p. 26-29)⁸.

Grande parte das considerações feitas até aqui dizem respeito ao modo como costumamos categorizar. Alguém poderia questionar qual a implicação dessas considerações em relação ao modo como deveríamos categorizar. Afinal, do fato de que normalmente não categorizamos empregando condições suficientes e necessárias, não se pode deduzir de maneira imediata que não devemos categorizar com base nessas condições.

⁷ Existem estudos dedicados a avaliar, em vários domínios conceituais, qual modelo é superior (*e.g.*, Estes, 1986)

⁸ Nesse sentido, outras concepções como a teoria-teoria de conceitos também são relevantes. De acordo a teoria-teoria, os conceitos mantêm entre si um relação semelhante àquela presente entre os termos de uma teoria científica (Margolis & Laurence, 2014). Para mais detalhes, ver Carey (2011).

De fato, a visão clássica pode ser útil em alguns casos. É bastante provável que certos conceitos bem definidos — como número par, quadrado, fêmea, avó, figura plana geométrica, *etc.* — possuam núcleos compostos por condições necessárias e suficientes, exigidas pela visão clássica. Em especial, no domínio das ciências há diversas definições que são elaboradas de forma bastante precisa, a partir da visão clássica, e posteriormente ensinadas a alunos que as utilizam para categorizar elementos. Em relação a conceitos surgidos dessa maneira, a presença de condições suficientes e necessárias pode servir como critério para justificar a categorização de algum elemento (Armstrong et al., 1983, p.263; Medin & Smith, 1984, p. 121-3).

No entanto, o problema é que grande parte dos conceitos não surge a partir de um processo de definição explícita desse tipo. Como esses conceitos não são criados a partir de condições suficientes e necessárias, mas a partir de diversas considerações, como, *e.g.*, avaliações de similaridades, pode simplesmente não haver um conjunto de condições suficientes e necessárias que sejam preenchidas por todos os indivíduos subsumidos por um dado conceito. Poderíamos delimitar um conceito artificialmente de acordo com a visão clássica, mas nesse caso, como observou Wittgenstein, sempre correríamos o risco de deixarmos de fora muitos dos exemplares normalmente associados ao conceito em questão.

Como os estudos sobre conceitos mencionados até aqui demonstram, há uma dificuldade enorme mesmo para explicar conceitos tão simples e concretos como o conceito de pássaro. Quando tentamos explicar ou definir conceitos mais abstratos essa dificuldade cresce exponencialmente (Armstrong et al., 1983, p. 305).

A falta de consenso acerca do conceito de moralidade é um exemplo disso. Inúmeros autores já tentaram definir o conceito de moralidade a partir da visão clássica e não obtiveram êxito. Há toda uma literatura filosófica dedicada a esse tema. Ainda na década de 1950, Alasdair MacIntyre iniciou um artigo com a seguinte frase: “The central task to which contemporary moral philosophers have addressed themselves is that of listing the distinctive characteristics of moral utterances” (1957, p.325). O artigo de MacIntyre foi republicado em 1970 em uma antologia chamada “The Definition of Morality” (Wallace & Walker, 1970), na qual estão reunidos artigos de diversos filósofos renomados como Elizabeth Anscombe, Kurt Baier, Philippa Foot, William Frankena e Peter Strawson. Todos eles empregaram argumentos sofisticados na tentativa de definir o conceito de moralidade, obviamente sem chegarem a um consenso. Até hoje trabalhos continuam sendo dedicados à tarefa de definir a moralidade sem que tenhamos qualquer perspectiva de convergência em torno de uma resposta (Stich, 2009, p. 220-1).

As sucessivas falhas nas tentativas de definir as condições necessárias e suficientes do conceito de moralidade sugerem que esse conceito segue o padrão de outros conceitos abstratos, como o de jogo. Desta forma, é bastante provável que essas supostas condições suficientes e necessárias, das quais supostamente teríamos um conhecimento tácito, simplesmente não existam (Stich, 2012, p. 92-4).

A moralidade, ao invés de possuir limites limpos e precisos, parece possuir várias zonas de penumbra e sobreposições com outros domínios. Em especial, em muitos casos é difícil distinguir claramente o campo/domínio da moral dos campos do direito, da religião e da etiqueta. No entanto, como a discussão sobre os problemas da visão tradicional deve ter deixado claro, esse não é um problema incomum ou fatal. O importante é que possamos caracterizar a moralidade de modo que pelo menos com relação a um considerável número de casos fique clara a sua distinção com respeito a outros conceitos, ainda que alguns casos fiquem em uma zona cinzenta (Warnock, 1971, p. 1-2).

No que diz respeito ao conceito de moralidade, não quero delimitá-lo de forma tão precisa que ele não corresponda a uma das formas normalmente utilizadas pelas pessoas em geral, mas também não quero deixá-lo tão borrado e tão aberto a ponto de as observações que farei nos próximos capítulos não corresponderem a ele em alguma medida.

Com relação ao conceito de inato, como veremos a seguir em uma seção específica, a dificuldade é ainda maior. Com relação a esse conceito não se trata apenas da impossibilidade de encontrar uma definição em termos de condições suficientes e necessárias. A confusão em torno do conceito de inato parece ser muito maior, o que me levará a questionar a sua própria utilidade.

2.1 Conceito de julgamento moral

Ainda que não seja possível elaborar uma definição tradicional para o conceito de moralidade que englobe todos os seus aspectos, é indispensável tentar esclarecer um pouco sobre o que estamos tratando. Como vimos, um dos obstáculos a serem enfrentados nesta pesquisa é a ausência de consenso entre filósofos, psicólogos, antropólogos e biólogos evolucionistas a respeito do significado de moralidade. Em muitos dos trabalhos que adotam uma perspectiva evolucionista, seja na área de filosofia ou de psicologia moral, não é fácil discernir sobre o que exatamente os autores estão falando. Essa é uma crítica que Stich (2009, p. 220-1) faz em relação a Prinz, mas que poderia ser estendida a outros autores.

A moralidade é um fenômeno bastante complexo que pode ser estudado sob vários ângulos. A partir de qualquer um desses ângulos, entretanto, é bastante difícil enxergar todos

aspectos envolvidos na moralidade. Assim, diferentes autores acabam enfatizando diferentes aspectos.

Há ainda a dificuldade de definir o conceito de moralidade decorrente da instabilidade das normas morais produzida pelas constantes mudanças históricas e sociais. Alguns autores adotam uma postura excessivamente estática e paroquial, assumindo a existência de uma espécie de paisagem moral congelada no tempo que aguarda ser descrita de forma definitiva, enquanto que o fenômeno que pretendem descrever, como demonstram os estudos antropológicos, se assemelha mais a um fluxo heraclitiano. No entanto, apesar das constantes mudanças, é possível que existam certas características mais abstratas que gozem de certa estabilidade. Essa estabilidade pode refletir-se na utilização bastante difundida de certos termos como ‘bom’, ‘mau’, ‘certo’, ‘errado’, ‘dever’ e alguns outros. Ou seja, embora os comportamentos que são condenados ou recomendados possam variar largamente, talvez a prática de qualificar condutas como boas ou más, certas ou erradas, seja um fenômeno bem mais constante, de caráter universal ou quase-universal entre seres humanos (Warnock, 1971, p. 2-10).

Outro obstáculo diz respeito à dificuldade de diferenciar o que é parte de uma definição e o que é parte de uma teoria moral normativa. Alguns, por exemplo, consideram que é parte da descrição do conceito de moralidade que ela serve apenas para disciplinar relações interpessoais. No entanto, outros podem considerar que a moralidade tem um escopo bem mais amplo. Assim, a afirmação mencionada não seria parte de uma descrição, mas de uma afirmação teórica substantiva acerca de como a moralidade deve ser. De modo semelhante, alguns afirmam que a moralidade se refere ao código de conduta que um indivíduo ou grupo consideram ser o mais importante (Hare, 1982, p.55-56). No entanto, outros podem questionar essa descrição, fazendo referência à importância que muitos indivíduos atribuem à religião (Gert & Gert, 2016).

Em razão de todos os obstáculos mencionados até aqui, não acredito ser possível encontrar a definição ‘correta’ de moralidade, no sentido de encontrar uma definição que englobe todos os usos habituais desse conceito. De qualquer maneira, para facilitar a compreensão dos capítulos seguintes, buscarei fixar algumas balizas para a utilização do conceito de moralidade. Estas balizas propostas a seguir não devem ser julgadas como corretas ou incorretas, mas como apropriadas ou inapropriadas (Whiteley, 1960, p. 141).

A delimitação que farei é necessária, pois em muitos casos o conceito de moralidade é fonte de ambiguidades. Em primeiro lugar, a palavra moralidade costuma ser utilizada em dois sentidos bastante amplos. No sentido descritivo, moralidade se refere a certos códigos de conduta aceitos por uma sociedade, por um grupo ou por um indivíduo como um guia para as suas condutas, ou seja, ela se refere ao modo como as pessoas, de fato, julgam as suas próprias

condutas e as de terceiros. Por sua vez, no sentido normativo, moralidade costuma se referir a um código de conduta hipotético que, dadas certas condições, seria aceito por todas pessoas racionais, ou seja, trata-se de uma idealização acerca de como as pessoas *deveriam* julgar condutas (Gert & Gert, 2016).

Na presente dissertação, a expressão ‘moralidade’ será utilizada em um sentido descritivo. Quando a utilizamos nesse sentido, ela pode se referir a códigos de condutas com conteúdos bastante distintos, de modo análogo ao que ocorre com o conceito de ‘direito’, o qual é aplicado, sem ambiguidade, em relação a sociedades que possuem normas muito distintas. Em determinado grupo, a moralidade pode permitir a escravidão ou conferir privilégios a certas pessoas, de acordo com o seu sexo, a sua etnia, a sua cor de pele, etc.; enquanto que, em outro grupo, pode proibir qualquer discriminação racial ou de gênero. O fato de aceitarmos uma certa descrição do conteúdo da moralidade não traz implicação direta para o modo como devemos nos comportar (Luco, 2014, p. 361-2; Gert & Gert, 2016).

Além de esclarecer que utilizarei o conceito de moralidade em um sentido descritivo, preciso ainda identificar a qual dos códigos de conduta esse conceito se refere. Até mesmo em sociedades pequenas e homogêneas é possível, por vezes, distinguir a moralidade de outros códigos de conduta; e, em sociedades mais complexas, essa distinção se torna bastante nítida (Gert & Gert, 2016).

A moralidade pode ser compreendida como um instituto paralelo ao direito, à etiqueta e à religião. Nas sociedades em que é possível fazer uma diferenciação entre essas instituições, ela costuma seguir a seguinte lógica. Como regra, a etiqueta diz respeito às normas que são consideradas menos sérias que as normas morais e que não são percebidas como normas objetivas, ou seja, as pessoas costumam ter a percepção de que elas dependem das opiniões e práticas sociais predominantes em certo momento⁹. Já o Direito se distingue da moralidade em razão de um grau maior de institucionalização, marcado pela presença de oficiais encarregados de interpretar as normas, usualmente explicitadas em textos, e aplicar as punições previstas. A Religião, por sua vez, distingue-se da moralidade na medida em que as suas normas de conduta são justificadas por meio de referências a eventos passados nos quais estão envolvidos seres supernaturais. Ainda que a Religião possa servir como fonte para as normas morais, muitas vezes podemos distingui-las, pois a Religião costuma envolver muito mais do que regras de conduta. Embora seja frequente a sobreposição das condutas proibidas ou exigidas pela religião

⁹ A distinção entre moralidade e etiqueta relaciona-se com a distinção entre normas morais e normas convencionais. Trataremos da oposição entre esses dois tipos de normas em uma seção específica.

e pela moralidade, a religião pode proibir ou exigir mais do que as normas morais (Gert & Gert, 2016; Bischof, 2012, p. 200-1).

Fazer referência à moralidade apenas como o código de conduta aceito pelo grupo ou pela sociedade causa ainda alguns outros problemas, pois as normas aceitas por um indivíduo do grupo não são exatamente as mesmas aceitas por outro. Esse problema se agrava muito no contexto atual, no qual existem sociedades extremamente grandes e heterogêneas, nas quais se torna difícil conceber algum código de conduta aceito igualmente por todos os indivíduos. Há, no entanto, a alternativa de analisar a moralidade a partir da perspectiva do indivíduo. Assim, a moralidade pode ser concebida, também, como as normas de conduta que um indivíduo aceita como guia de suas condutas (Luco, 2014, p. 361-2).

Na presente dissertação, o nosso objetivo é compreender como a capacidade de fazer julgamentos morais se desenvolve nos indivíduos e de que maneira essa capacidade está relacionada às mudanças ocorridas ao longo da evolução de nossa espécie. Nesse sentido, o nosso objeto de estudo são os sistemas cognitivos existentes no indivíduo e relacionados à moralidade (I-moralidade). A moralidade enquanto conjunto de normas e valores existentes em uma sociedade (E-moralidade) interessa-nos apenas secundariamente, na medida em que ela influencia a conformação da I-moralidade.

Faço esta distinção com base naquela proposta por Chomsky entre *E-language* e *I-language* (Chomsky, 1986). Chomsky afirma que o foco de uma teoria linguística fundada sobre a biologia deve ser os sistemas cognitivos presentes dentro do indivíduo, a língua interna (*I-language*). Em contraste com essa língua interna, existem as línguas enquanto objeto de estudos linguísticos históricos, ou seja, as línguas externas (Inglês, Espanhol, Francês, etc.) resultantes das propriedades da população de falantes (Fitch, 2010, p. 32-34).

A distinção que faço se assemelha àquela feita por Whiteley entre um conceito *sociológico* e um conceito *psicológico* de moralidade. No sentido sociológico, moralidade diz respeito aos padrões de comportamentos que são ensinados e recomendados aos membros da sociedade, enquanto, no sentido psicológico, diz respeito ao conteúdo da consciência do próprio indivíduo, às condutas que ele acredita que devem ser adotadas por ele e pelos outros membros do grupo (Whiteley, 1960, p. 141-2).

O conceito de moralidade apropriado para a presente pesquisa é esse conceito psicológico (I-moralidade). Estou interessado na moralidade, principalmente, sob a perspectiva do indivíduo e de seus julgamentos acerca do que é certo e do que é errado. Nesse sentido, o conceito de moralidade é aqui relevante à medida que estiver atrelado à ideia de julgamento

moral. Os esclarecimentos que farei sobre a moralidade servirão para delimitar o espaço no qual os julgamentos morais são feitos.

Utilizarei a expressão ‘julgamento moral’, de modo semelhante ao utilizado por Mameli (2013, p. 903-4), para fazer referência a certo tipo de estado mental tipicamente expresso por afirmações com a seguinte estrutura: A ação ou omissão X, dadas as circunstâncias e as pessoas envolvidas, é moralmente obrigatória, proibida ou permitida. Tendo em vista que ‘moralmente obrigatório’, ‘moralmente proibido’ e ‘moralmente permitido’ são interdefiníveis, podemos simplificar dizendo que os julgamentos morais seguem a forma ‘X é moralmente obrigatório’¹⁰.

A noção de obrigação moral envolve a demanda para que uma ação seja praticada ou para que ela não seja praticada (Mackie, 1990, p.33) e vincula-se diretamente à ideia de que quem não a cumpre merece algum tipo de punição. Essas demandas podem ser expressas por meio de imperativos do tipo ‘mantenha as suas promessas!’ ou ‘não trabalhe aos sábados!’, etc. (Mameli, 2013, p. 904-5).

Embora a moralidade envolva fenômenos no nível populacional, ou seja, embora ela não ocorra apenas na cabeça dos indivíduos, esse tipo de representação mental sobre o que é moralmente obrigatório parece ser central para a moralidade. Vários filósofos, como Bernard Williams (2006, p. 174-5), já enfatizaram a relevância da noção de obrigação moral e a sua relação com a demanda por determinada ação. Williams trata a moralidade como uma forma particular de uma categoria mais ampla, a ética. Ao adotar essa posição, adoto um conceito mais estrito de moralidade. Um conceito mais amplo não restringiria moralidade apenas a obrigações, pois muitas das coisas que recebem o adjetivo ‘moral’ transcendem essa noção de obrigação. Há muitas condutas que atraem admiração – *e.g.*, dedicar uma vida inteira a ajudar outras pessoas — e podem ser rotuladas como ‘morais’, em um sentido amplo, mas que não envolvem a execução de qualquer obrigação (Skorupski, 1993, p. 124). No entanto, como me interessa particularmente os aspectos psicológicos envolvidos na produção dos julgamentos morais, é prudente adotar o conceito mais estrito.

Em suma, na presente dissertação utilizo o conceito de moralidade em um sentido descritivo, a fim de fazer referência, como regra, a certas normas de conduta aceitas por um indivíduo e utilizadas por ele para julgar se as suas condutas e aquelas de outros indivíduos estão em conformidade com as obrigações que lhes são cabíveis. Essa caracterização preliminar ainda não é suficiente para distinguir as normas morais de outras normas. A fim de tornar essa

¹⁰ O ato X é moralmente proibido se e somente se não X for obrigatório, e X é permissível se e somente se não for moralmente obrigatório não X (Mameli, 2013, p. 903).

delimitação mais precisa, discutirei a seguir alguns dos critérios tradicionalmente empregados por filósofos para definir a moralidade.

2.1.1 *Características formais e materiais da moralidade*

Há filósofos, como Hare (1982), que tentam distinguir a moralidade de outros códigos de conduta por meio de critérios formais. No entanto, filósofos como Bernard Williams (1993, p. 73-74), Warnock (1971) e Frankena (1966, p. 691-2) consideram que esses critérios são insuficientes para delimitar a moralidade. Segundo eles, qualquer delimitação significativa envolveria também referências ao *conteúdo* da moralidade.

A seguir veremos quais são os critérios formais normalmente utilizados para delimitar a moralidade, quais são as críticas feitas à utilização deles e, por fim, examinaremos algumas propostas que procuram suprir essas supostas falhas por meio da inclusão de critérios materiais.

Predominância. Hare propôs que para uma orientação de conduta ser qualificada como moral, além de ser universalizável e prescritiva, ela deve ser tratada como predominante (*overriding*) em relação aos outros tipos de orientação (Hare, 1982, p.55-56). Esse critério parte da ideia de que algumas vezes dois preceitos entram em conflito, no sentido de ser impossível agir de acordo com ambos. Em casos assim, nós permitimos que um preceito se sobreponha ao outro. Para Hare, isso não envolve apenas agir de um modo e não de outro, mas também pensar que devemos agir de certo modo, ainda que isso implique desconsiderar o outro preceito. Hare não considera que um princípio, para ser qualificado como moral, precisa *ser* predominante; ele apenas diz que ele deve *ser tratado* como predominante (Hare, 1982, p. 56; Hare, 1977, p. 168). Desta forma, o simples fato de um preceito moral não se sobrepor a outro tipo de preceito em um caso concreto não faz com ele deixe de ser moral.

No entanto, até mesmo Hare não considera que ser predominante é uma condição necessária para que um preceito seja considerado moral. Ele admite que em alguns casos outros princípios podem se sobrepor aos princípios morais. Segundo ele, os princípios morais se dividiriam em duas subclasses: (1) princípios morais que são tratados sempre como predominantes; (2) princípios morais que, embora não sejam sempre predominantes, são selecionados por meio de uma forma de pensamento crítico que se vale de referências aos princípios morais da primeira subclasse. Portanto, se quiséssemos saber se alguém trata algum princípio como um princípio moral, deveríamos perguntar a essa pessoa se ela, em alguma circunstância, deixaria que esse princípio fosse afastado (*overriden*) em nome de algum outro. Se ela responder que não, então ela o está tratando como um princípio moral. Mas ainda que ela

responda sim, é possível que se trate de um princípio moral do segundo tipo (Hare, 1982, p. 60-1).

A crítica frequentemente feita em relação a esse critério diz respeito à possibilidade de considerações estéticas ou religiosas serem consideradas mais importantes do que as considerações que normalmente qualificamos como morais, tais como evitar causar danos a terceiros (Gert & Gert, 2016).

Hare considera o termo ‘moral’ ambíguo e não busca definir um único sentido correto de moralidade. O objetivo dele é estabelecer critérios para distinguir a moralidade — compreendida em um sentido que seria o relevante para as discussões travadas entre os filósofos morais — de outros domínios. Portanto, simplesmente apontar exemplos da utilização do termo moralidade que não se enquadram nos critérios propostos por Hare não é suficiente para refutar a utilização desses critérios.

Como não estou buscando encontrar uma definição de moralidade no sentido da teoria clássica de conceitos, não é necessário entrar nas minúcias envolvidas na justificação dos princípios considerados morais, mas que não são sempre predominantes em relação aos demais. Como busco apenas delinear uma espécie de *cluster concept* da moralidade, o que é importante observar é que em grande parte dos casos as normas morais apresentam essa característica da predominância.

Independência de autoridade. De acordo com esse critério, a moralidade não seria considerada contingente em relação aos desejos de ninguém. A sua força normativa não seria derivada de sua aprovação por alguma autoridade específica, por mais prestigiosa ou poderosa que fosse. Assim, um princípio moral não perderia a sua força caso descobríssemos que não é endossado por algum indivíduo ou alguma entidade (antepassados, deuses, *etc.*) considerada importante na comunidade (Mameli, 2013, p. 909).

Esse critério, juntamente com o critério da inescapabilidade que veremos a seguir, é sintetizado por Mackie, quando ele diz que o julgamento moral “involves a call for action, or for the refraining from action, and one that is absolute, not contingent upon any desire, or preference or policy or choice, his own or anyone else’s” (Mackie, 1990, p. 33).

A independência em relação à autoridade também não é um critério essencial da moralidade, como demonstra uma pesquisa conduzida por Kelly e Stich (Kelly *et al.*, 2007). Eles elaboraram um estudo conduzido pela Internet acerca de certas práticas danosas envolvendo casos paradigmáticos de violações de normas morais. Nessa pesquisa, os entrevistados foram questionados sobre uma série de transgressões (Kelly *et al.*, 2007). Foi proposto, por exemplo, um cenário no qual havia uma lei proibindo professores de bater nos

alunos. Determinada professora conhecia a proibição e ainda recebeu instruções de seu diretor para não bater nos alunos. No entanto, quando um aluno por diversas vezes tumultuou a aula e espancou os seus colegas, a professora deu-lhe algumas palmadas. No segundo cenário, não havia nenhuma lei contra professores baterem em alunos e o diretor disse que a professora poderia bater nos alunos que se comportassem mal, se ela quisesse. Os entrevistados tiveram de responder se havia algum problema em a professora dar palmadas no aluno que se comportou mal. Quase metade dos entrevistados entendeu que a professora não estava errada no segundo cenário, enquanto apenas 8% interpretaram desta forma no primeiro cenário. Ou seja, para uma parte substancial daqueles que responderam à entrevista, normas proibindo castigos físicos são dependentes de alguma autoridade que as instituem. Em cinco outros casos, resultados semelhantes foram obtidos.

Universalidade. De acordo com o critério da universalidade, quem diz que alguma ação é moralmente certa ou errada estaria disposto a adotar essa mesma posição em relação a qualquer outra situação que fosse semelhante. Um julgamento moral seria um preceito universalizável, no sentido de que o proponente desse preceito estaria disposto a aplicá-lo tanto a si mesmo quanto a qualquer outro em situações futuras (Mackie, 1990, p. 83-8).

Até aqui deve ter ficado claro que a moralidade envolve a avaliação de ações. Portanto, não é qualquer evento da natureza que está sujeito a ser julgado moralmente. Mas é preciso também qualificar quais são os possíveis agentes envolvidos nessas ações. Filósofos costumam dizer que para uma norma ser considerada moral ela deve ser aplicável às ações de todos os agentes morais e, em seguida, costumam definir o significado de ‘agente moral’ (Taylor, 1978, p. 42-3). Warnock, *e.g.*, considera que os julgamentos morais são direcionados a todos os “seres racionais.” Para ele, a condição necessária para que um agente seja apto a ser julgado moralmente não é propriamente o fato de pertencer à espécie humana, mas o fato de ser racional. Segundo ele, um agente deveria ser considerado racional quando dispusesse de alguma habilidade para perceber e levar em consideração as diferentes ações possíveis para atingir certos fins; avaliar o que pode ser dito a favor e contra essas alternativas; escolher alguma delas; e, por fim, agir de acordo com a sua escolha (Warnock, 1971, p. 14).

A existência de diversos relatos históricos sobre julgamentos de condutas de animais (Srivasta, 2007) torna questionável a afirmação de que ser racional, no sentido estipulado por Warnock, é uma condição necessária para um agente ser julgado moralmente. É possível interpretar a posição de Warnock de forma caridosa e considerar que a condição necessária é, na verdade, que quem julgue considere que o agente da ação a ser julgada é racional, ainda que de fato ele não seja.

Dessa forma, alguém que julgue moralmente uma porca por assassinato poderia estar imaginando que ela dispõe das competências elencadas por Warnock. Os estudos de Viveiros de Castro sobre o perspectivismo ameríndio indicam que em certas teorias indígenas é comum a prática de se atribuir a outras espécies predicados tipicamente atribuídos aos seres humanos, tais como o de ser racional. De acordo com essa visão, “os animais são gente, ou se veem como pessoas.” Cada espécie seria “um mero envelope (uma ‘roupa’) a esconder uma forma interna humana, normalmente visível apenas aos olhos da própria espécie ou de certos seres transespecíficos, como os xamãs” (Castro, 1996, p. 117).

No entanto, ainda assim, não teríamos como saber ao certo se todos que julgam moralmente de fato consideram, em seus julgamentos, que o agente da ação analisada preenche os requisitos de racionalidade mencionados. De qualquer maneira, Warnock captura uma característica interessante dos julgamentos morais, ainda que talvez ela não seja necessária para a existência desses julgamentos: quem julga moralmente alguma ação costuma considerar que o agente escolheu agir daquela forma e que ele poderia ter escolhido agir de modo diverso, ou seja, que ele dispunha de uma alguma forma de livre arbítrio.

Se, por um lado, é possível que os julgamentos morais sejam direcionados a outras espécies, por outro, o fenômeno do etnocentrismo faz com que muitas vezes indivíduos da espécie humana não sejam considerados dignos de serem julgados moralmente. Em algumas sociedades tradicionais, os julgamentos morais são direcionados apenas àqueles provenientes das mesmas origens dos demais membros da tribo. É comum que nas línguas faladas nessas sociedades a palavra equivalente a ‘homem’, ‘ser humano’ ou ‘pessoa’ sirva apenas para designar os membros do grupo. Em outros tantos casos, os membros das tribos vizinhas são considerados seres humanos, mas de um tipo degenerado: eles são bárbaros, irracionais, grosseiros, abomináveis, *etc.* (Bischof, 2012, p. 140).

Há ainda outros estudos que tornam questionável a aplicação do critério da universalidade para delimitar o domínio da moralidade. Em um deles, conduzido por Kelly e Stich, foram propostos alguns cenários para testar a universalidade de normas morais que proíbem condutas danosas. Num dos estudos, 52% dos participantes entenderam que há 300 anos não havia problemas em castigar fisicamente um marinheiro que estava bêbado em serviço, enquanto apenas 6% entenderam que tal punição poderia ser aplicada nos dias de hoje. Em outro caso, uma parte muito maior dos entrevistados considerou ser a escravidão moralmente mais aceitável na Grécia antiga ou em Roma do que atualmente (Kelly & Stich, 2007, p. 362-4).

Há evidências de que muitas vezes o sentimento de raiva em relação a violações morais só é despertado quando o violador pertence à mesma comunidade daquele que julga a ação.

Como observa Mameli, devido à globalização e aos ideais universalistas difundidos no Ocidente, hoje em dia algumas pessoas consideram todos os seres humanos, ainda que vivam em regiões distantes do planeta, como se fossem membros de sua própria comunidade. Em razão disso, eles direcionam a sua raiva em relação a qualquer um que viole normas consideradas morais. Se eles julgam que a escravidão é moralmente condenável, eles julgam que ela era condenável também na Antiguidade, quando costumava ser aceita. A prática de estender os julgamentos morais dessa forma não é universal e era rara no passado, quando em geral as pessoas não se enxergavam como pertencentes a uma comunidade moral composta por todos os seres humanos ou por todos os seres racionais (Mameli, 2013, p. 914-5).

Desta forma, em vez de falarmos de universalidade, seria mais apropriado dizermos que os julgamentos morais são extensivos aos demais membros da comunidade relevante. A extensão dessa comunidade relevante dependerá do contexto cultural específico no qual o julgamento é feito. Em certos contextos, ela poderá abranger todos os seres humanos ou todos os seres racionais. O importante aqui é destacar que os julgamentos morais não são feitos para serem aplicáveis a apenas um caso específico. Quem julga moralmente, como regra, pretende que o seu julgamento seja aplicável também à conduta de outros indivíduos em uma situação semelhante.

Inescapabilidade: imperativos hipotéticos e não-hipotéticos. As normas morais são concebidas como inescapáveis (Joyce, 2006; Foot, 1972; Williams 2006). Considerar uma demanda prática como inescapável significa considerá-la aplicável aos indivíduos independentemente dos objetivos, desejos e preferências que possuam. Desta maneira, quando fazemos um julgamento moral, considerações sobre se a ação praticada seria boa para o violador, sob o ponto de vista prudencial, são irrelevantes. Assim, por exemplo, a nossa indignação em relação a um assassino não diminui quando aprendemos que ele auferiu benefícios financeiros com o assassinato (Mameli, 2013, p. 908). Quem está moralmente obrigado a adotar alguma conduta não se desobriga de praticá-la ao demonstrar que ela não favorece os seus desejos ou interesses, tampouco em razão de preferir não estar vinculado ao sistema de normas morais vigentes. As obrigações morais se aplicam às pessoas mesmo que elas não queiram (Williams, 2006, p. 177-8). Em outras palavras, podemos dizer que as normas morais são imperativos não hipotéticos.

Kant classificava os imperativos em hipotéticos e categóricos. Imperativos hipotéticos apresentam a necessidade prática (*praktische Notwendigkeit*) de se adotar determinada ação para obter algum objetivo desejado. Eles dizem apenas que uma determinada ação é boa para alcançar algum objetivo. Imagine, por exemplo, que você está em uma rodoviária e encontra um

viajante chamado Pedro. Por algum motivo, você acredita que Pedro quer ir para São Paulo e diz, apontando para um ônibus que está prestes a partir: ‘Pedro, você deve pegar aquele ônibus!’ Se você descobrir que Pedro mudou de ideia e não quer mais ir para São Paulo, a frase que você pronunciou perderá a sua força; não haverá mais qualquer suporte que dê sustentação ao ‘dever’ de entrar no ônibus.

Por sua vez, o imperativo categórico apresenta uma determinada ação que é objetivamente necessária (*objektiv-notwendig*). Ele diz que uma ação deve ser tomada por si própria, independentemente de qualquer objetivo ou intenção, ou seja, ela não deve ser concebida como meio para alguma outra finalidade (1999, p. 36-7). De acordo com a doutrina kantiana, seguida nesse ponto por muitos outros filósofos, as normas morais são exemplos paradigmáticos de imperativos categóricos (Williams, 2006, p. 177-8; Mackie, 1990, p. 29; Taylor, 1978, p. 45; McDowell & McFetridge, 1978, p. 13).

De fato, quando expressamos um comando com palavras como ‘dever’ em um contexto moral, não precisamos fundamentá-lo por meio de referências aos interesses ou desejos do destinatário desse comando. As normas morais não são concebidas como imperativos hipotéticos ou prudenciais, mas como orientações que devemos seguir independentemente dos nossos desejos e interesses, *i.e.*, elas possuem o que Sripada e Stich chamam de *independent normativity* (Sripada & Stich, 2007, p. 284-5).

No entanto, não nos parece apropriado qualificar as normas morais como imperativos categóricos pelo simples fato de elas não serem contingentes em relação aos desejos de seus destinatários. Há outros tipos de imperativos não hipotéticos aos quais seria estranho atribuir a denominação de ‘imperativo categórico.’ As normas de etiqueta, *e.g.*, não deixam de ser aplicáveis a alguém que não tenha interesse em segui-las ou que não se importe com elas. Sob a perspectiva da etiqueta, *devemos* mastigar de boca fechada, ainda que prefiramos mastigar de boca aberta. Algo semelhante ocorre em relação às normas de um clube. Imagine que a secretária diga a um dos sócios que ele não *deve* fumar nas dependências do clube. Se ele responder que renunciará ao seu título de sócio na próxima semana e que não se importa com a sua reputação, ela não dirá algo como: ‘desculpe-me, eu estava errada, o senhor pode continuar fumando.’ Em casos assim, o dever não deixa de estar fundamentado pelo simples fato de não ter conexão com os desejos do agente; portanto, ele é utilizado aqui também de modo não hipotético. Tanto as normas de etiqueta quanto as normas morais são inescapáveis no sentido de que um comportamento não deixa de ofendê-las em razão do agente ser indiferente a elas. Assim, se utilizássemos a classificação binária proposta por Kant, ao constatarmos que esses

imperativos não são hipotéticos teríamos de concluir que eles são categóricos (Foot, 1972, p. 307-9)¹¹.

Contudo, aqueles que defendem a visão ortodoxa de que as normas morais são imperativos categóricos insistiriam que as normas de etiqueta são, na verdade, imperativos hipotéticos. Eles diriam que o agente ao qual essas normas se destinam poderia questionar, sem incorrer em irracionalidade, se ele possui alguma razão para orientar as suas condutas de acordo com essas normas; enquanto que em relação às normas morais esse tipo de questionamento não seria possível (Foot, 1972, p. 307-9). Esses filósofos se apegam à ideia kantiana de que o simples fato de percebermos a existência de uma norma moral automaticamente nos daria uma razão para agir. McDowell, por exemplo, diz que um agente virtuoso é motivado pela razão a se comportar de acordo com as demandas da moralidade e que essa motivação não é condicionada à posse de qualquer tipo de desejo. Ou seja, um agente virtuoso não é caracterizado como alguém que tem certos desejos e sentimentos considerados nobres, mas como alguém com certas crenças sobre o mundo. A diferença entre alguém virtuoso e alguém não virtuoso seria uma diferença de percepção e de cognição: quem enxerga o mundo como o homem virtuoso sobrepõe imediatamente as demandas da moralidade sobre todas as outras (McDowell & McFetridge, 1978, p. 31-2).

A ideia de que normas morais fornecem razões para agir mesmo a agentes que não se importam nem um pouco com as consequências do descumprimento dessas normas parece ser uma simples ilusão, resultante de uma tentativa de conferir força mágica ao dever moral (Foot, 1972, p. 314). Mas não pretendo discutir aqui a posição ontológica dos imperativos categóricos. Na presente dissertação, abordo a moralidade pela perspectiva de quem faz julgamentos morais. Dessa forma, não é algo muito relevante sabermos se imperativos categóricos existem de forma objetiva, ou mesmo se a ideia de imperativo categórico faz algum sentido¹², mas sim sabermos se quem julga moralmente concebe as normas morais como imperativos categóricos.

Nessa perspectiva, não parece correto classificar como julgamentos morais apenas aqueles feitos como base em supostos imperativos categóricos. É possível que muitos, de fato, concebiam as normas morais dessa forma, mas, para o propósito da presente pesquisa, não convém restringir de tal maneira o conceito de moralidade.

¹¹ De acordo com Foot demandas morais são hipotéticas no sentido de que a influência que elas exercem no agente é condicionada pela presença de certos desejos, como viver abertamente, ser respeitado pelos seus vizinhos, etc. (Foot, 1972, p. 314; McDowell & McFetridge, 1978, p. 13). No entanto, estamos aqui abordando a moralidade pela perspectiva de quem emite julgamentos morais e, como tentamos demonstrar, quem emite tais julgamentos não costuma concebê-los como hipotéticos.

¹² Foot questiona se afirmações como “a truly moral man acts ‘out of respect for the moral law’” ou “he does what is morally right because it is morally right” fazem algum sentido (Foot, 1972, p. 310-313).

Podemos imaginar alguém que julgue uma ação com base nas normas de sua sociedade e não as conceba como imperativos categóricos, mas apenas como regras que facilitam a convivência e que, para que sejam mantidas, precisam ter o seu cumprimento fiscalizado pelo maior número possível de pessoas. Não me parece razoável dizer que esses julgamentos não são julgamentos morais simplesmente porque essa pessoa não os concebe como imperativos categóricos.

Em suma, considerarei que os julgamentos morais envolvem imperativos não hipotéticos, mas ignorarei a discussão sobre se eles, conceitualmente, envolvem ou não imperativos categóricos.

Comportamento pró-social: requisito material. Como mencionamos anteriormente, diversos filósofos consideram os critérios formais insuficientes para delimitar a moralidade (Williams, 1993, p. 73-74; Warnock, 1971; Frankena, 1966, p. 691-2). Segundo Frankena, *e.g.*, o conceito de moralidade resultante de critérios formais, em razão de ser muito amplo, é insuficiente para distinguir as demandas práticas decorrentes de normas morais daquelas de demandas estéticas. Imagine alguém que considere que a sua própria vida é uma espécie de obra de arte e que em todas as suas decisões práticas busque apenas encontrar a melhor forma de completá-la. Essa pessoa pode orientar o seu comportamento e os seus julgamentos acerca do comportamento de outras pessoas apenas por critérios estéticos. A maneira como ela concebe esses critérios pode preencher todos os requisitos formais mencionados, ou seja, ela pode conceber as normas estéticas como universais, inescapáveis, predominantes sobre quaisquer outras e existentes independentemente da vontade de qualquer autoridade específica.

Frankena propõe então, como uma tentativa para solucionar esse problema, o acréscimo de um critério material aos critérios formais, de modo a restringir um pouco o conceito de moralidade. De acordo com esse critério material, um preceito moral:

includes or consists of judgments (rules, principles, ideals, etc.) that pronounce actions and agents to be right, wrong, good, bad, etc., simply because of the effect they have on the feelings, interests, ideals, etc., of other persons or centers of sentient experience, actual or hypothetical (or perhaps simply because of their effects on humanity, whether in his own person or in that of another). Here 'other' may mean 'some other' or 'all other.' On this conception, a morality must embody some kind of social concern or consideration; it cannot be purely prudential or purely aesthetic (Frankena, 1966, p. 688-9).

Com esse critério, Frankena pretende deixar bastante espaço para diferentes tipos de moralidade, tais como as deontológicas, as utilitárias, as nacionalistas (nazismo, fascismo), as baseadas em estratificação social, etc. No entanto, ele pretende excluir do domínio da moralidade considerações que sejam puramente egoístas ou estéticas (1966, p. 691-2).

Para Frankena a função da moralidade – com a palavra função ele quer dizer a razão pela qual uma prática existe e é mantida – não é apenas servir como um guia supremo para as condutas, mas tornar possível a cooperação e as atividades sociais entre os seres humanos (Frankena, 1966, p. 691).

Outros filósofos, como Warnock, adotam posição semelhante. Para ele a moralidade serve para melhorar a difícil situação na qual os seres humanos vivem — “the amelioration of the human predicament” (1971, p. 16-7). Warnock parte da constatação de que os recursos, assim como as informações, as nossas habilidades e a nossa inteligência, são limitados. Muitas vezes não agimos racionalmente na resolução de nossos problemas e nas nossas relações. Além disso, frequentemente estamos em competição com outras pessoas e somos vulneráveis aos seus ataques, ao mesmo tempo em que dependemos delas para nossa sobrevivência. Por fim, a nossa empatia por outros seres humanos é limitada. Todas essas condições tornam difícil a tarefa de garantir a cooperação em um grau necessário para que consigamos suprir as nossas necessidades básicas (1971, p. 23). A função da moralidade seria justamente contrabalançar essas dificuldades, potencializando a empatia e, assim, reduzindo a produção de danos (1971, p. 26).

Como regra, as pessoas costumam se preocupar mais em satisfazer os seus próprios interesses do que em auxiliar as outras (Mackie, 1990, p. 108). As normas morais acabam formando um sistema que freia essa tendência por meio da proteção de interesses distintos aos do próprio agente (Mackie, 1990, p. 106; Bischof, 2012, p. 202).

Como vimos na introdução, a ideia de que a função primordial da moralidade é reprimir o egoísmo e incentivar a cooperação serve como base para os estudos conduzidos sob a égide da ética evolucionista descritiva. A ênfase na relação entre moralidade e comportamento pró-social está presente de maneira muito forte nos estudos contemporâneos destinados a explicar a moralidade a partir de uma perspectiva naturalista (*e.g.*, Haidt, 2012, cap. 11).

Esse destaque dado à cooperação e à repressão do egoísmo acaba levando muitos autores a classificarem como morais apenas as normas que se destinam a evitar a produção de danos ou lesões a terceiros, tais como as normas que proíbem assassinatos, mutilações, condutas que causem sofrimento, *etc.* Um dos critérios utilizados por Turiel, *e.g.*, para distinguir as normas morais das convencionais era o tipo de justificção oferecida para as normas. Segundo ele, as normas morais eram justificadas com base nos danos causados a terceiros ou à justiça (Turiel, 1983; Tisak & Turiel, 1984).

No entanto, diferentes etnografias demonstram que a moralidade envolve muito mais do que apenas normas sobre condutas lesivas. Como veremos no capítulo destinado especificamente à teoria dos domínios morais, Haidt emprega vários desses estudos para

demonstrar que a moralidade não se restringe a condutas que causam danos ou ferem alguma noção de justiça. Muitas sociedades não ocidentais possuem uma moral muito mais abrangente: elas moralizam práticas, tais como tabus alimentares, que não têm qualquer relação com a produção de dor ou de danos a pessoas (Haidt, 2012).

Ainda assim, até mesmo as normas que não se destinam a evitar lesões a pessoas também servem de alguma forma para reprimir o egoísmo. Essa ideia está presente no critério proposto por Frankena, quando ele diz que as normas morais não protegem apenas os interesses de indivíduos reais, mas abrangem também a proteção de entes hipotéticos. Nesse sentido, as normas que visam proteger alguma ordem sagrada ou alguma ideia de pureza também poderiam ser vistas como formas de reprimir o egoísmo, em nome de algo considerado importante na comunidade.

De qualquer maneira, o incentivo a comportamentos pró-sociais é inegavelmente uma das características marcantes das normas morais, ainda que eventualmente seja possível identificar algumas normas morais que não cumpram diretamente essa função.

Em suma, os julgamentos morais costumam apresentar as seguintes características que os distinguem de outros julgamentos normativos: eles são predominantes; a sua força normativa não depende da aprovação de alguma autoridade específica; são universalizáveis ou, ao menos, extensíveis aos demais membros da comunidade relevante; são inescapáveis no sentido de sua aplicabilidade não depender dos objetivos do agente ao qual eles se destinam; e servem para promover comportamentos pró-sociais ou para frear tendências egoístas em nome de algo considerado importante por outros indivíduos.

2.1.2 Normas morais e normas convencionais

Um argumento frequentemente utilizado para defender a ideia de que a moralidade é fruto da evolução baseia-se na pesquisa feita por Elliot Turiel sobre a distinção entre normas morais e convencionais (Dwyer, 2007, p. 239–242; Joyce, 2006, p. 134–137; Hauser, 2007, p. 291; Machery & Mallon, 2010, p. 32). Turiel e seus colegas afirmam que, desde muito cedo e em todas as culturas, as crianças são capazes de distinguir entre esses dois tipos de normas (Turiel, 1983; Tisak & Turiel, 1984). Crianças reconhecem que regras sobre vestimentas, alimentos e outros aspectos da vida são convenções sociais, *i.e.*, arbitrárias e modificáveis em certos aspectos. Um exemplo de regra desse tipo seria a obrigação de utilizar uniforme nos colégios. Por outro lado, elas compreendem que regras que proíbem ações que machucam outras pessoas são regras morais — *e.g.*, empurrar outras crianças para fora do balanço ou puxar seu cabelo.

Embora as crianças demonstrem sensibilidade a essa distinção desde muito cedo — entre dois e três anos de idade —, as informações que elas recebem não seriam suficientes para explicar a capacidade para distinguir entre esses dois tipos de normas, pois os pais não articulariam verbalmente a distinção e as puniriam pelas violações em ambos os casos, muitas vezes com equivalente severidade (Dwyer, 2008, p. 416-7). O limitado número de transgressões praticadas ou testemunhadas pelas crianças (*impoverished moral input*) não seria suficiente para explicar a ampla capacidade que elas possuem para fazer julgamentos em relação a casos com os quais não estão familiarizadas (Dwyer, Huebner & Hauser, 2010, p. 492). Além disso, as transgressões de normas convencionais seriam muito mais numerosas do que as transgressões de normas morais, e também seriam corrigidas com uma frequência muito maior pelas pessoas responsáveis por cuidar de crianças. Desta forma, as crianças teriam poucas oportunidades de observar o tipo de reação emocional peculiar associado à transgressão de normas morais (Dwyer, 2008, p. 416). Dwyer afirma ser difícil, mesmo para adultos, distinguir entre reações emocionais fortes: não é fácil saber se o interlocutor está com raiva, irritado, desapontado com uma conduta, indignado ou enojado. É pouco provável que crianças sejam melhores em fazer esse tipo de discriminação sutil em relação às respostas das pessoas responsáveis por cuidar delas (Dwyer, 2008, p. 416-7).

Prinz, no entanto, contesta esta afirmação. Segundo ele, ela carece de comprovação empírica, uma vez que existem pesquisas que indicam que crianças e adultos são bons em distinguir emoções (*e.g.*, Field & Walden, 1982). Além do mais, a diferença no reforço das normas não seria devida apenas a diferenças nas emoções apresentadas pelos educadores. Há outras diferenças como a indução de empatia, comum nos casos de normas contra comportamentos que causam danos (Prinz, 2008, p. 433-4).

Turiel originalmente propôs os seguintes critérios para distinguir as normas morais das convencionais. As normas morais seriam: (a) concebidas como dotadas de força prescritiva objetiva, independente da autoridade de qualquer indivíduo ou instituição; (b) entendidas como aplicáveis universalmente, ou seja, não seriam aplicáveis apenas localmente em determinado momento, mas prescreveriam comportamentos para outras culturas e para outros contextos históricos; (c) justificadas com base nos danos causados a terceiros ou à justiça; e (d) a violação desse tipo de norma seria considerada mais séria. Por outro lado, normas convencionais seriam aquelas dependentes de alguma autoridade, aplicáveis localmente, justificadas com base em convenções e cuja violação é considerada menos séria (Turiel, 1983; Tisak & Turiel, 1984; Nichols, 2005, p. 356-7; 2008, p. 260; Kelly & Stich, 2007, p. 353).

A distinção entre normas morais e convencionais estaria presente em diferentes faixas etárias, desde crianças com menos de quatro anos até adultos, e nas mais diversas culturas, religiões e classes sociais (Kelly & Stich, 2007, p. 354). Entre os grupos culturais e religiosos pesquisados estão crianças chinesas em idade pré-escolar (Yau & Smetana, 2003), crianças coreanas (Song *et al.*, 1987), nigerianas (Holloos *et al.*, 1986), e provenientes das ilhas Virgens, adolescentes e adultos (Nucci *et al.*, 1983), crianças amish e menonitas, crianças e adolescentes integrantes da Igreja Reformada Holandesa (Nucci & Turiel, 1993).

Apesar dessas pesquisas, a distinção proposta por Turiel é suscetível a muitas críticas. Autores que a enfatizam tendem a considerar as normas sobre condutas lesivas (*harm norms*) como o aspecto central da moralidade. Nichols, por exemplo, ao elaborar o conceito de *core moral judgment* enfatiza o papel que emoções associadas a condutas lesivas desempenham na capacidade para distinguir entre normas morais e convencionais (2008, p. 265).

No entanto, diferentes etnografias, tais como aquelas feitas por Richard Shweder, demonstram que a moralidade envolve muito mais do que apenas normas sobre condutas lesivas. Em especial, um estudo conduzido por ele (Shweder, Mahapatra & Miller, 1987), na Índia, no estado de Orissa, ilustra o caráter etnocêntrico, voltado para culturas individualistas, da teoria de Turiel. Nesse estudo, os entrevistados tinham de responder questionários sobre pequenas histórias nas quais alguma regra era violada. A grande maioria dos entrevistados indianos respondeu que violações de regras referentes à comida, ao sexo e às vestimentas feriam a moralidade, não convenções sociais. Praticamente não havia indícios de um pensamento baseado em normas convencionais nessa sociedade. Como regra geral, as normas eram percebidas como universalmente válidas e inalteráveis, a ordem social se confundia com a ordem moral. Certamente, em muitas outras culturas, práticas homossexuais, bem como a transgressão de inúmeros outros tabus relacionados à pureza e à santidade seriam consideradas transgressões morais. Como salienta Haidt, a distinção entre moral e convenção não é universal, e a sua existência e intensidade relacionam-se ao caráter individualista ou sociocêntrico de determinada cultura:

The distinction between morals and mere conventions [...] turns out to be a cultural artifact, a necessary by-product of the individualistic answer to the question of how individuals and groups relate. When you put individuals first, before society, then any rule or social practice that limits personal freedom can be questioned. If it doesn't protect somebody from harm, then it can't be morally justified. It's just a social convention (Haidt, 2012).

De fato, os estudos apresentados pelos defensores da distinção moral-convencional não consideram uma série de populações autóctones. Em especial, eles não consideram grupos que vivem de acordo com condições semelhantes àquelas que os humanos enfrentavam no período

do Pleistoceno, o qual foi essencial no processo de conformação da mente humana. Em muitas culturas, a distinção entre normas convencionais e morais não parece ser aplicável. Em sociedades tradicionais, regras referentes à dieta e aos ornamentos são compreendidas como morais. As pessoas que violam esse tipo de norma são castigadas e, frequentemente, apresentam sentimentos de vergonha ou de culpa (Prinz, 2008, p. 385). Em certas sociedades autóctones não se concebe a possibilidade de alterar normas sociais por meio de convenções; todas as normas obedecidas são fruto da tradição e a única forma possível de mudança é pela reinterpretação da tradição.

Grande parte dos estudos iniciais que confirmavam a distinção moral-convencional foram feitos em sociedades individualistas, e aqueles feitos em culturas predominantemente sociocêntricas estavam evitados de um sério problema em sua elaboração: eram realizados em situações muito limitadas e envolviam basicamente questões sobre práticas danosas — do tipo que crianças nas primeiras séries do colégio poderiam cometer — contrapostas a questões sobre normas convencionais não baseadas em fortes reações de nojo e repulsa. Estudos mais amplos com adultos recomendam ceticismo a respeito dos critérios propostos por Turiel (Kelly *et al.*, 2007; Gabennesch, 1990; Prinz, 2008).

As pessoas julgam algumas ações sérias e independentes de qualquer autoridade, mas não as justificam com apelos a danos causados, seja à justiça ou a direitos — *e.g.*, masturbar-se com a carcaça de uma galinha morta antes de comê-la, limpar o banheiro com a bandeira nacional ou manter relações sexuais consentidas com um irmão utilizando dois métodos contraceptivos diferentes (Haidt *et al.*, 1993). Não oferecer uma parte do último pedaço de bolo para quem estiver presente viola uma norma moral, mas não é algo muito sério. Andar nu, por outro lado, viola uma convenção, mas pode ser considerado grave. Além disso, muitas vezes as pessoas relutam em generalizar para outras culturas normas que são justificadas em razão dos danos causados por seu descumprimento (Machery & Mallon, 2010, p. 34; Prinz, 2008, p. 384; Prinz, 2008, p. 430). No mesmo sentido, pesquisas feitas por Nichols (2002a; 2002b; 2004) indicam que estudantes americanos consideram até mesmo certas normas de etiqueta, especialmente aquelas que proíbem condutas consideradas nojentas (*disgust-inducing*), transgressões sérias e independentes de qualquer autoridade. Esses resultados demonstram como transgressões que não envolvem a produção de danos, lesões de direitos ou violações de justiça também despertam reações que seriam típicas à violação de normas morais (Kelly & Stich, 2007, p. 361).

Além disso, há evidências de que os pais não respondem da mesma maneira a violações de normas que consideram morais. As crianças podem aprender a diferenciar normas morais e

convencionais imitando e internalizando os diferentes padrões de raciocínio empregados pelos educadores. Há evidências de que os pais adaptam os seus estilos de intervenção disciplinar conforme o tipo de regra violada. Regras morais seriam reforçadas por afirmação de poder e apelo a direitos; regras convencionais, por raciocínio e apelos à ordem social¹³ (Prinz, 2008, p. 392-3).

Por fim, normas destinadas a evitar a produção de danos e lesões também podem ser consideradas como não universais e dependentes de autoridade em certos contextos, como demonstram os já mencionados estudos sobre a proibição da escravidão e de castigos físicos a alunos (Kelly *et al.*, 2007; Kelly & Stich, 2007, p. 362-4).

Em suma, muitas evidências indicam que a pretensão de universalidade das teorias inspiradas por Turiel sobre a distinção entre normas morais e convencionais é infundada, e que elas não são adequadas sequer para descrever como os indivíduos de sociedades ocidentais compreendem as normas em geral. Talvez a distinção possa revelar apenas um viés psicológico no sentido da universalização das normas que consideramos mais importantes. Ainda que a distinção entre normas morais e convencionais esteja presente em algumas culturas apenas, e não da maneira proposta pelos que a enfatizam, ela pode indicar um importante fator que diferencia as normas consideradas morais das demais: a presença de certas emoções. A distinção pode servir para mostrar como um tipo muito específico de cultura criou as condições necessárias para que algumas normas fossem desprovidas de uma parte de seu conteúdo emocional e passassem a ser consideradas normas convencionais. No entanto, a distinção entre normas morais e convencionais não serve como indício da existência de alguma forma de conhecimento moral anterior ao aprendizado.

2.2 Conceito de inato

Além das ressalvas feitas quanto ao conceito de moralidade, não devemos também ignorar que o termo ‘inato’ costuma ser empregado de diversas maneiras, causando grande confusão no debate sobre o nativismo moral.

Mameli e Bateson realizaram um estudo sobre como o conceito de inato é utilizado na literatura científica. Nesse estudo eles encontraram 26 definições diferentes que buscavam identificar o conceito de inato com uma única propriedade biológica ou psicológica. Dentre essas definições, eles selecionaram oito “finalistas” que poderiam ser cientificamente úteis – as definições restantes eram apenas derivações imperfeitas das finalistas. No entanto, nenhuma

¹³ Prinz apresenta detalhadamente exemplos que indicam como normas concebidas como morais são reforçadas de maneira distinta (2008, p. 431-2).

delas foi considerada satisfatória (Mameli & Bateson, 2006; Mameli & Bateson, 2011, p. 436): cada uma dessas definições conseguia capturar alguns dos aspectos presentes no senso comum acerca do conceito de inato, mas deixava outros de fora.

Uma forma comum de caracterizar os traços inatos é pela oposição aos traços adquiridos. No entanto, há várias noções diferentes de ‘aquisição’ e não está claro qual noção é relevante para a compreensão do significado de inato. Por exemplo, se considerarmos que uma característica é adquirida se e somente se há um período no qual o organismo em questão a possui e um momento anterior no qual ele não a possui, concluiremos que todas as características fenotípicas são adquiridas, tendo em vista que nenhuma delas está presente no zigoto (Samuel, 2004, p. 136). Outra alternativa seria fixar um critério temporal, tal como o momento do nascimento. Nesse sentido, uma característica seria adquirida caso a sua aquisição ocorresse após o nascimento. Se a aquisição ocorresse antes do nascimento, ela seria inata. No entanto, essa classificação não é satisfatória, tendo em vista a existência de aprendizado pré-natal. Crianças aprendem, antes mesmo de nascer, a reconhecer a voz de sua mãe e são também capazes de reconhecer as histórias que ouvem ainda no útero (Decasper & Spence, 1986; Spence & Decasper, 1987), mas ninguém diria que elas têm um conhecimento inato sobre contos infantis. Além disso, há traços considerados inatos, como a presença de pelos pubianos e de outros traços sexuais secundários que só se manifestam (são adquiridos) anos após o nascimento (Samuels, 2004, p. 137).

A alternativa de substituir a dicotomia inato versus adquirido pela dicotomia inato versus aprendido também é insatisfatória, pois muitos dos traços que se desenvolvem sem aprendizado não são considerados inatos. Indivíduos que sofrem danos no córtex pré-frontal podem apresentar o que Damásio (2008) chama de sociopatia adquirida, um tipo de transtorno de personalidade que não é considerado inato e nem resulta de processos de aprendizado (Mameli & Bateson, 2011, p. 437).

A fonte das confusões é o fato de as inúmeras definições de inato não serem coextensivas, de modo que uma característica considerada inata de acordo com uma definição pode não ser considerada inata quando adotamos outra (Mameli & Bateson, 2006, p. 177-178). Esse fato pode indicar que o conceito de inato confunde propriedades que talvez devessem ser mantidas separadas sob o ponto de vista das ciências biológicas e cognitivas. Mameli e Bateson oferecem o exemplo do termo ‘jade’ para ilustrar essa hipótese (*clutter hypothesis*). Pensava-se que esse termo correspondia a um único composto químico. No entanto, descobriu-se que ele era utilizado para fazer referência a dois compostos distintos, nefrita e jadeíta (Mameli & Bateson, 2011, p. 436 e 440).

Os diferentes conceitos de inato analisados por Mameli e Bateson correspondem a certos aspectos do conceito vernacular de inato, e muitas vezes uma característica classificada como inata conforme um desses conceitos acaba também sendo considerada inata quando adotamos outro dos conceitos elencados. Essa sobreposição, entretanto, não é perfeita; um traço que é inato de acordo com um conceito pode não ser inato de acordo com outro. Quando alguém utiliza o termo ‘inato’, costuma ter em mente uma combinação de diferentes características. Essa prática de utilizar o termo ‘inato’ para fazer referência a diferentes propriedades que não se sobrepõem perfeitamente conduz a uma série de confusões. Um tipo bastante comum de confusão consiste em demonstrar que uma característica possui determinada propriedade e, a partir daí, simplesmente inferir que ela também possui outras propriedades usualmente relacionadas com características inatas (Mameli & Bateson, 2011, p. 440).

Griffiths, Machery e Linqvist defendem a ideia de que o conceito de inato é apenas uma expressão da crença presente no senso comum de que os seres vivos possuem uma natureza (Griffiths, Machery & Linqvist, 2009). De acordo com a visão de leigos a respeito da biologia (*folkbiology*), algumas características do organismo decorrem de uma natureza interna enquanto outras são impostas pelo ambiente.

Com o desenvolvimento da genética, o senso comum passou a identificar essa natureza interna com o genoma. Algumas características seriam determinadas pelos nossos genes, enquanto outras decorreriam da influência do ambiente (Mameli & Bateson, 2006, p. 155; Griffiths, 2011, p. 320-5; Samuels, 2004, p. 138). Contudo, essa compreensão é falsa. Embora a diferença entre dois indivíduos possa ser causada por diferenças genéticas ou por diferenças no ambiente de desenvolvimento, o desenvolvimento de qualquer traço depende tanto dos genes quanto do ambiente. Hoje há um consenso de que qualquer traço é resultado da interação dos genes com outros fatores ambientais — *interactionist consensus* (Kitcher, 2001).

Todas as formas de vida conhecidas dependem da existência de um conjunto de genes e de condições ambientais que favoreçam o desenvolvimento do organismo (luz, oxigênio, temperatura adequada, *etc.*). Nos casos de organismos multicelulares como os seres humanos, o desenvolvimento depende ainda de interações entre as células e o ambiente no qual elas estão localizadas, isto é, o interior do organismo que está em desenvolvimento. Esse raciocínio pode ser aplicado a qualquer traço, desde a presença de uma mão ou de um pulmão até a manifestação de um traço comportamental como a linguagem ou a moral (Fitch, 2010, p. 27-28; Richerson & Boyd, 2005, p. 9-10). Um feto não desenvolve qualquer de seus traços — braços, pernas, cérebro, *etc.* – se não houver trocas de nutrientes, de água e de oxigênio com a sua mãe. Da mesma forma, após o nascimento, o desenvolvimento de qualquer traço dependerá de interações

com fatores externos ao organismo (ingestão de alimentos, temperatura, pressão atmosférica e gravidade adequadas, *etc.*).

Muitos nativistas, entretanto, costumam considerar que esses fatores ambientais funcionam apenas como desencadeadores (*triggering*) de processos previamente estabelecidos geneticamente (Samuels, 2004, p. 137). Diante da constatação de que o ambiente é essencial para o desenvolvimento de qualquer traço, costuma-se apelar à ideia que o ambiente fornece apenas as condições necessárias para que os traços inatos se desenvolvam, mas não contribuem com informações a respeito de como esse traço se desenvolve. O genoma seria uma espécie de projeto arquitetônico contendo as informações a respeito do organismo, enquanto os demais fatores ambientais seriam apenas os insumos necessários para a execução desse projeto¹⁴.

De acordo com essa visão, aceita por muitos psicólogos e biólogos, traços inatos seriam aqueles geneticamente codificados. Essa visão é atrativa, pois ela se coaduna com a intuição de que os traços inatos devem, de alguma maneira, estar presentes desde o começo da vida do organismo: embora os traços ainda não tenham se manifestado, eles já estariam lá de forma codificada (Mameli & Bateson, 2011, p. 438).

A utilização de discursos que assumem a premissa de que genes codificam o desenvolvimento de determinados traços é bastante comum – gene da miopia, da hemofilia, da obesidade *etc.* Esses discursos consideram que genes codificam determinados traços que se manifestam quando o organismo se desenvolve em ambientes normais (*standard*). No entanto, esse discurso pode ser invertido. Poderíamos dizer com a mesma legitimidade que uma paisagem natural codifica o desenvolvimento desses mesmos traços, quando as condições genéticas padrão estão presentes (Griffiths & Gray, 1994, p. 283-4).

Dependendo do sentido adotado para o conceito de informação, podemos sim dizer que os genes ‘carregam informação’ a respeito de fenótipos, mas nenhuma das noções de informação disponíveis é suficiente para fazer uma distinção entre os papéis do genoma e dos demais recursos ambientais. De acordo com a noção matemática de informação derivada da teoria da comunicação (Johnston, 1987), um evento transmite informação sobre outro na medida em que está correlacionado com ele (covariação). Ocorre que qualquer traço está correlacionado tanto com fatores genéticos quanto não genéticos. Assim, a adoção desse conceito implica um tratamento simétrico dos diferentes fatores causais. Poderíamos dizer que os fatores não genéticos envolvidos no desenvolvimento formam o canal pelo qual as informações contidas nos genes são transmitidas. No entanto, poderíamos também inverter os papéis de transmissor

¹⁴ Esse tipo de visão sobre as características inatas se baseia em uma divisão nítida entre *nature* e *nurture*. Abrantes (2014b) explora as possibilidades de superação dessa dicotomia.

(*sender*) e canal de transmissão. Ou seja, é igualmente correto tratar os genes como o canal que permite a transmissão das informações contidas nos fatores não genéticos (Griffiths & Gray, 1994, p. 281-2; Samuels, 2004, p. 137). O seguinte exemplo ilustra como também podemos interpretar que o ambiente codifica o desenvolvimento do fenótipo:

Consider, for example, the case of the development of normal species-typical human limbs. According to folk intuitions, this trait is innate. The maternal womb is one of the environmental factors required for the development of this phenotype. Hence, the maternal womb should be a mere supportive factor for the development of species typical human limbs. But changes in the maternal womb affect the structure of the developing limbs. For example, a thalidomide-bathed womb generates the development of limbs that differ markedly from the species-typical limbs that develop in thalidomide-free wombs (...) More generally, it is likely that for every phenotype some environmental factor can be found such that changes in this environmental factor produce changes in the phenotype (rather than just the death of the organism) (Mameli & Bateson, 2006, p. 162-3).

A *Figura 1* demonstra como a presença de talidomina pode influenciar o desenvolvimento dos membros em seres humanos. Assim como ocorre com braços e pernas, qualquer dos traços usualmente considerados inatos pode sofrer algum tipo de modificação em razão da alteração de algum fator ambiental.



Figura 1 – Foto de Louise Medus com seu pai (Courtenay-Smith, 2008)

Quem afirma que os genes codificam o desenvolvimento de certos traços, costuma adotar, ainda que de maneira inconsciente, um conceito intencional ou semântico de informação. Conforme esse tipo de conceito, quando as condições do canal de transmissão são alteradas e os genes não carregam a informação esperada acerca do fenótipo, os genes são considerados mal interpretados (Griffiths & Gray, 1994, p. 281-2). No entanto, esse tipo de noção só faria algum sentido se considerássemos que há um arquiteto por trás da formação dos genes, de maneira que quando o resultado final não correspondesse à intenção do arquiteto haveria uma falha de comunicação¹⁵. Obviamente, esse tipo de pensamento não é consistente com a premissa, amplamente aceita, de que as mutações que dão origem a novos genes resultam de um processo

¹⁵ Alguns biólogos renomados, entretanto, como Ernst Mayr (1974), procuraram defender a utilização da linguagem teleológica contra esse tipo de objeção.

de variação cega. Em suma, a ideia de que genes codificam certos traços¹⁶ não serve para sustentar a distinção entre traços inatos e não inatos.

Há diversas outras tentativas de definir o conceito de inato elaboradas a partir da concepção de que os seres possuem uma natureza interna. Griffiths identificou as principais qualidades que costumam ser atribuídas aos traços decorrentes dessa suposta natureza:

Fixity: The trait is hard to change; its development is insensitive to environmental inputs in development; its development appears goal-directed or resistant to perturbation.

Typicality: The trait is part of what it is to be an organism of that kind; every individual has it, or every individual that is not malformed, or every individual of a certain age, sex, or other natural subcategory.

Teleology: This is how the organism is meant to develop; to lack the innate trait is to be malformed; environments that disrupt the development of this trait are themselves abnormal (Griffiths, 2011, p. 321).

A crença na existência de uma natureza dos seres permeou as discussões científicas, assumindo, entretanto, uma nova roupagem por meio da utilização de conceitos próprios da Biologia. No lugar de fixidez, emprega-se a ideia de *canalização*. Uma característica qualquer é canalizada quando ela é relativamente insensível à manipulação de parâmetros ambientais (Griffiths, 2011, p. 323). Em relação à qualidade de *tipicidade*, é feita a ressalva de que ela não se refere à universalidade de uma característica, mas ao fato de ela ser *típica da espécie* - *species-typical* (Griffiths, 2011, p. 323). Por fim, o substituto biológico óbvio para a ideia de teleologia é o conceito de adaptação no sentido darwiniano. No entanto, aqui há uma diferença importante. Uma adaptação não deve ser compreendida em um sentido teleológico, embora seja comum dizermos que um indivíduo apresenta má-formação quando não desenvolve um fenótipo que foi moldado pela seleção natural (Griffiths, 2011, p. 323). Essa maneira de descrever o desenvolvimento do indivíduo que não apresenta determinada adaptação típica da espécie (por exemplo, um ser humano que não possui olhos) é bastante frequente, mas ela não se coaduna com a teoria da evolução, tendo em vista que a seleção natural não se relaciona à existência de um *telos*¹⁷. A seguir, veremos os problemas apresentados pelas tentativas já feitas de equiparar ‘inato’ a alguma dessas propriedades.

Andre Ariew publicou uma série de artigos (1996; 1999; 2007) nos quais defendeu a ideia de que o conceito de inato pode ser definido por meio da noção de ‘canalização ambiental.’ O conceito de canalização foi introduzido pelo biólogo Conrad H. Waddington (1942). Waddington definiu canalização como “the capacity [of development] to produce a particular

¹⁶ Como diz Peter Godfrey-Smith: “All the genes can code for, if they code for anything, is the primary structure (amino acid sequence) of a protein” (Godfrey-Smith, 1999, p. 328).

¹⁷ Essa questão não é totalmente pacífica na Biologia. James Lennox, por exemplo, acredita que Darwin concebia as explicações selecionistas sobre as adaptações como explicações teleológicas (Lennox, 1993).

definite end-result in spite of a certain variability both in the initial situation from which development starts and in the conditions met during its course” (Waddington 1975, p. 99). Waddington ilustrou a sua ideia com a famosa paisagem de desenvolvimento (*Figura 2*).

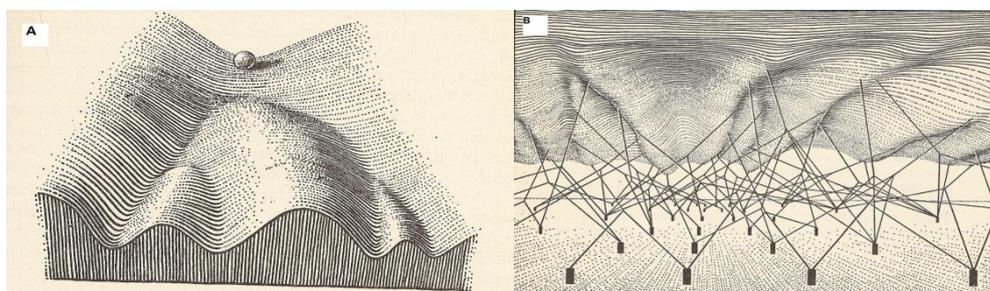


Figura 2 – Paisagem de desenvolvimento (Griffiths, 2009)

A canalização pode ser ambiental ou genética. No caso desta última, diferentes organismos chegam no mesmo ponto final de desenvolvimento a partir de pontos de partida genéticos diferentes; ou seja, o fenótipo em questão é protegido (*buffered*) contra variações genéticas. No caso de canalização ambiental, diferentes organismos atingem o mesmo ponto final de desenvolvimento apesar de interagirem com diferentes fatores ambientais. Em suma, a canalização é ambiental caso o traço fenotípico em questão seja relativamente insensível à manipulação de parâmetros ambientais; ou genética, caso ele seja relativamente insensível à manipulação de parâmetros genéticos (Griffiths, 2011, p. 323; Griffiths, 2009; Mameli & Bateson, 2011, p. 439-40).

Assim, conforme Ariew, os traços inatos seriam aqueles com alta canalização ambiental¹⁸. De acordo com essa definição, o ‘nativismo’ de um fenótipo seria uma questão de grau, pois não há qualquer forma absoluta de canalização. Sempre haverá certas condições ambientais que tornarão o desenvolvimento de uma característica inviável. Portanto, um traço será mais inato quanto maior for o número de parâmetros ambientais contra os quais o desenvolvimento estiver protegido (Griffiths, 2009). A presença de pernas em seres humanos, por exemplo, seria em grande medida inata, uma vez que o desenvolvimento delas em indivíduos com genótipos semelhantes é bastante insensível às condições ambientais (Samuels, 2004, p. 138).

A identificação do conceito de inato com o conceito de canalização ambiental satisfaz a intuição de que os traços inatos são aqueles cujo desenvolvimento é estável; ou seja, cujo desenvolvimento é difícil de ser evitado em situações normais (Mameli & Bateson, 2011, p. 439-40; Samuels, 2004, p. 138). Há, de fato, variação entre os diferentes traços quanto à sua

¹⁸ Sober (1998) adota posição semelhante: “a phenotypic trait is innate for a given genotype if and only if that phenotype will emerge in all of a range of developmental environments.”

sensibilidade aos fatores ambientais. Alguns traços, como a presença de cinco dedos em cada mão nos seres humanos, são pouco afetados pelo espectro normal de variações ambientais. Já outros traços, como o tamanho do corpo, costumam variar bastante, mesmo entre indivíduos geneticamente semelhantes, em razão de fatores como a alimentação durante a infância (Richerson & Boyd, 2005, p. 9-10).

Contudo, a insensibilidade a variações ambientais não é algo muito fácil de ser medido. A nossa capacidade visual, por exemplo, é imune a diversas variações ambientais. Ela se desenvolve tanto em indivíduos criados no Kalahari quanto naqueles criados na Floresta Amazônica ou no Círculo Polar Ártico. No entanto, basta alterar uma simples variável, a presença de luz, e essa capacidade não se desenvolverá. Nesse sentido, a ideia de que um traço é robusto em relação às variações ambientais tem algo de circunstancial. Ela significa que o traço é robusto em relação a certo número de variações ambientais que ocorrem comumente, no presente momento histórico, nos locais onde o organismo em questão está presente. Podemos imaginar que em outro momento e em outro local variações que impedem a presença da característica em questão poderiam ser comuns.

Outra crítica feita a esse conceito de inato apoia-se no fato de que, de acordo com ele, um traço pode ser simultaneamente inato e aprendido. Existem traços que apresentam baixíssima variação, mas apenas porque as condições ambientais exigidas para o aprendizado são onipresentes. Por exemplo, quase todo ser humano adquire a crença de que o sol nasce todos os dias, e o desenvolvimento dessa crença é altamente canalizado, no sentido de ser insensível às variações ambientais presentes nos diversos ambientes ocupados pelo homem (Samuels, 2004, p. 138). Como veremos, Sterelny (2003) salienta como processos de construção de nichos podem fazer com que o desenvolvimento de traços aprendidos possua alto grau de canalização ambiental. Segundo ele, esta seria a explicação para várias das capacidades cognitivas humanas, ou seja, elas seriam ao mesmo tempo aprendidas e canalizadas (Mameli & Bateson, 2011, p. 439-40).

Nos diversos debates nativistas, é também bastante comum a alegação de que uma característica é inata em razão de ela desempenhar alguma função. A relação entre adaptação e natureza humana é ainda mais estreita quando analisamos os estudos sobre essa concepção presente no senso comum. Pessoas leigas correlacionam de forma muito estreita o fato de uma característica desempenhar uma função e de ela “estar no DNA” (Linguist *et al.*, 2011). Muitas vezes o que se discute nos debates nativistas é o grau de especificidade de determinada capacidade, ou seja, se ela resulta de mecanismos gerais de aprendizado ou se é fruto de um mecanismo específico a ela dedicado, como, *e.g.*, um módulo mental (Fitch, 2010, p. 86). A

partir desse tipo de discussão desliza-se facilmente para discussões adaptacionistas, pois a existência de um mecanismo cognitivo destinado exclusivamente à cognição moral, *e.g.*, seria um forte indício de que esse tipo de cognição é uma adaptação darwiniana (Sterelny, 2010, p. 280).

Um dos problemas em identificar o conceito de inato com o conceito de adaptação é que um traço pode ser aprendido e, ainda sim, ser uma adaptação darwiniana. Dessa forma, um traço poderia ser ao mesmo tempo inato e aprendido. Costuma-se supor que adaptações são traços cujo desenvolvimento não envolve aprendizado, mas isso não é verdade. Tanto os traços que são aprendidos como aqueles não aprendidos podem se espalhar em uma população em razão da atuação da seleção natural. Mameli e Bateson citam como o exemplo a habilidade desenvolvida pela águia-pesqueira para obter alimentos:

Fish-catching is a species-typical behaviour pattern of the osprey, an eagle that uniquely snatches with its talons healthy fish from the water's surface. This skill emerges in ospreys as a result of a long and complicated learning process. Is this behaviour a Darwinian adaptation? Suppose that at some stage the ospreys didn't exhibit this behaviour at all. Then a genetic mutation G that allows or motivates ospreys to learn to catch fish from the water appeared in some individual. Fish-catching was fitness enhancing: fish-catchers had higher fitness (on average) than non-catchers. Fish-catching was heritable because G was genetically transmissible, and natural selection increased the frequency of fish-catchers and decreased the frequency of non-catchers. On this scenario, fish-catching is a standard Darwinian adaptation of the ospreys, despite being a learned trait (Mameli & Bateson, 2006, p. 174).

Várias das habilidades humanas e de outros mamíferos e aves seguem esse mesmo padrão, tal como a nossa habilidade de caminhar sobre duas pernas (Mameli & Bateson, 2011, p. 438-9).

Outro problema é que, de acordo com essa visão, as doenças genéticas e quaisquer novas mutações genéticas não seriam consideradas inatas, pois elas não são produtos da seleção natural (Mameli & Bateson, 2011, p. 438-9).

Por fim, a presença de uma característica em praticamente todos os membros da espécie é frequentemente interpretada, até mesmo por alguns cientistas, como uma prova de que ela é inata. No entanto, a aplicação desse tipo de raciocínio conduz a conclusões contraintuitivas. Por exemplo, crenças compartilhadas universalmente, como a crença de que água mata a sede, teriam de ser consideradas inatas. De modo semelhante, a presença de uma flora intestinal, deveria ser interpretada como uma característica inata (Griffiths, 2009).

A *folk biology* costuma confundir essas três propriedades. Contudo, determinado traço pode ser uma adaptação, mas demandar um ambiente específico para se manifestar. Por exemplo, se um ambiente específico estava disponível de maneira estável no período em que a moralidade evoluiu, poderíamos esperar que mesmo sendo uma adaptação ela não emergisse

frente a grandes perturbações ambientais. Por outro lado, a moralidade poderia ser um efeito secundário (*by-product*) de outras adaptações e desenvolver-se de forma robusta diante de grandes variações (Joyce, 2013, p. 533). De modo semelhante, um traço pode ser uma adaptação e não ser típico da espécie. Algumas adaptações são monomórficas – todos os seres humanos perderam a habilidade para sintetizar a vitamina C; já outras são polimórficas – alguns adultos são capazes de metabolizar a lactose enquanto outros não (Griffiths, 2011, p. 324). A capacidade de metabolizar a lactose é uma adaptação, mas não é típica da espécie, tendo em vista que a maioria dos seres humanos adultos não a possui.

Alguns utilizam o termo ‘inato’ no sentido de que a característica em discussão apresenta essas três propriedades; outros utilizam esse termo tendo apenas uma delas em mente, ou ainda diferentes combinações. Essa ambiguidade do termo serve, por sua vez, para acobertar certas falácias como, por exemplo, inferir que uma característica é típica da espécie com base na constatação de que ela é uma adaptação. Se esse tipo de inferência fosse feita de forma explícita, ficaria claro que a conclusão não se segue da premissa. No entanto, quando a discussão é conduzida por meio da utilização do termo inato, é possível utilizá-lo em diferentes partes do texto para fazer referência a propriedades biológicas distintas e deslizar facilmente de uma propriedade para outra (Linguist *et al.*, 2011, p. 446).

A descrição que Konrad Lorenz fazia dos traços inatos é um exemplo desse tipo de confusão. Lorenz considerava que um traço era inato quando era uma adaptação e desenvolvia-se independentemente das variações ambientais (Griffiths & Gray, 1994, p. 279-80). No entanto, em uma famosa crítica ao conceito de inato utilizado por Lorenz, Daniel S. Lehrman demonstrou que não há uma conexão necessária entre esses dois elementos: uma determinada característica pode ser uma adaptação darwiniana e ser sensível a variações em fatores ambientais. Lehrman mencionou diversos exemplos, tais como o comportamento das ratas em relação aos seus filhotes: elas só constroem ninhos e abrigam os seus filhotes neles se forem expostas a certas variações de temperatura e se tiverem a chance de carregar objetos em suas bocas no começo de suas vidas (Lehrman, p. 342-3, 1953).

A linha de raciocínio seguida pelos psicólogos evolucionistas também costuma refletir confusão semelhante. Eles assumem que se uma característica é uma adaptação então ela é também canalizada em relação a mudanças ambientais e, portanto, irá desenvolver-se no ambiente atual ainda que tenha deixado de ser adaptativa. Essa hipótese é a base para o chamado ‘*adaptive thinking*.’ No caso da evolução humana, esse tipo de pensamento costuma ser empregado da seguinte forma. Em primeiro lugar, tenta-se descobrir como era o ambiente no Pleistoceno, época em que as principais características cognitivas humanas teriam supostamente

evoluído. Em seguida, busca-se definir quais estruturas cognitivas promoveriam um acréscimo na aptidão dos indivíduos inseridos nesse ambiente. Por fim, conclui-se que estas estruturas são agora típicas da espécie humana e se desenvolvem em todos os indivíduos normais. Há, contudo, uma falha nesse raciocínio: ainda que considerássemos que não houve mudanças significativas em nossos genes desde o Pleistoceno, por que deveríamos assumir que os genes correlacionados com o desenvolvimento de estruturas cognitivas bem adaptadas ao ambiente daquela época continuariam a produzir, nos dias de hoje, essas mesmas estruturas cognitivas, apesar das grandes diferenças ambientais entre os dois períodos? A hipótese de que as adaptações cognitivas surgidas durante o Pleistoceno são canalizadas em relação às mudanças ambientais ocorridas desde então depende de comprovação por meio de evidências empíricas que, até o momento, não foram apresentadas (Mameli & Bateson, 2006, p. 180-1).

Diante da constatação da impossibilidade de definir o conceito de inato por meio da referência a uma única propriedade, resta a alternativa de defender a hipótese de que esse conceito se refere a um *cluster concept*, isto é, a um conjunto de propriedades que tendem a ocorrer de forma concomitante. De acordo com essa hipótese, o conceito de inato é utilizado para fazer referência a algo real, uma espécie de ‘síndrome’ presente em alguns traços, aqueles inatos, e ausente em outros. Logo, as controvérsias entre nativistas e antinativistas poderiam ser interpretadas como desavenças acerca de quais os traços apresentam essa síndrome (Mameli & Bateson, 2011, p. 441).

A fim de que essa hipótese possa ser aceita, é indispensável que as diferentes propriedades usualmente atribuídas ao conceito de inato possuam uma forte correlação, mas esse não parece ser o caso. Como procurei demonstrar até aqui, é comum que diferentes pessoas associem uma propriedade ao conceito de inato, enquanto outras pessoas associam uma propriedade bastante diferente. Até o momento, não há evidências de que exista o que Wittgenstein chamava de *semelhanças de família* entre os diversos traços que são classificados como inatos. Enquanto essas evidências não forem apresentadas, a utilização desse conceito não apresenta qualquer utilidade e é uma fonte inesgotável de confusão (Griffiths, 2009).

Muitos autores sugerem que a distinção inato versus adquirido não tem qualquer valor científico (Linguist *et al.*, 2011, p. 451). Seguindo essa visão, propõem que a categoria ‘inato’ seja eliminada do discurso científico. Os vários debates nativistas deveriam ser abordados de forma diferente, por meio de referências específicas às propriedades que estão sendo discutidas. Desta forma, ficaria mais claro quais são as questões empíricas que estão em discussão. No lugar de utilizar o conceito de inato, que não tem um sentido definido, e, portanto, conduz facilmente a ambiguidades, é melhor fazer referência às propriedades específicas que estão

sendo discutidas, ainda que isso possa dar mais trabalho e ocupar mais espaço (Bateson 1991, p. 22).

Ainda que o conceito de inato não seja útil sob o ponto de vista teórico, isso não significa que o debate nativista em torno da moralidade não tenha qualquer sentido. A percepção das ambiguidades envolvidas na utilização desse conceito pode servir para o aperfeiçoamento da abordagem adotada para tratar de debates desse tipo. Há muitas questões importantes envolvidas que podem ser melhor respondidas se utilizarmos conceitos mais claros. Mameli e Bateson explicitam algumas das questões frequentemente envolvidas:

Some nativist debates are about whether the development of a particular trait involves learning or some other mechanism for adaptive plasticity. Others are about whether a trait can be modified by learning or by some other kind of environmental intervention, independently of the way it developed. Some nativist debates are really about whether a particular trait is environmentally canalized with respect to a specific range of variation; they are about whether we can affect the development of a trait via interventions that are available to us. Some debates are about the evolutionary origins of phenotypes, while others are about the broad heritability of a specific trait in a specific population. And so on (Mameli & Bateson, 2011, p. 442).

No âmbito da discussão sobre o nativismo moral, o conceito de inato é utilizado com todos os diferentes sentidos que mencionamos ao longo do presente capítulo. No entanto, Joyce acredita que as definições de inato utilizadas nesse debate costumam se agrupar em dois grupos principais: concepção *evolutiva* (*evolutionary conception*) e concepção *desenvolvimental* (*developmental conception*). De acordo com o primeiro grupo, ‘nativismo moral’ equivale à afirmação de que a moralidade é uma adaptação no sentido darwiniano — Joyce (2006, pos. 54) e Prinz (2008, p. 370) podem ser enquadrados nesse grupo. Por outro lado, aqueles que adotam a concepção desenvolvimental consideram um traço inato quando o seu surgimento está protegido (canalizado) contra variações no ambiente onde ocorre o desenvolvimento do organismo — Sripada (2008, p. 323-4), *e.g.*, adota esse tipo de concepção (Joyce, 2013, p. 532-33).

O conceito de inato manifesta-se também nas discussões sobre nativismo moral na forma da distinção entre características inatas e aprendidas. Muitas vezes os autores que defendem que uma determinada característica é inata procuram demonstrar que ela não poderia ter sido aprendida. Esse é o caminho adotado pelo famoso argumento da pobreza de estímulos utilizado por Noam Chomsky para defender que os princípios da Gramática Universal são inatos¹⁹. A inferência presente nesse argumento segue esta lógica: como os princípios de sintaxe não poderiam ter sido aprendidos, logo eles são inatos. Nesse mesmo sentido, as diversas experiências que procuram demonstrar o desenvolvimento precoce da moralidade pretendem, no

¹⁹ O argumento da pobreza de estímulos será explicado de maneira mais detalhada no capítulo sobre a Gramática Moral Universal.

fundo, demonstrar a presença de elementos que não foram aprendidos. Os antinativistas, por sua vez, procuram demonstrar que Chomsky está errado ao supor que o conhecimento desses princípios não poderia ser aprendido, mas não questionam a inferência de que, se não foi aprendido, então é inato (Mameli & Bateson, 2011, p. 437).

O conceito de inato utilizado por Haidt também repousa sobre essa dicotomia. De acordo com Haidt, ‘inato’ significa “organizado antes de qualquer experiência” (Haidt, 2012, cap. 12). Haidt concebe o cérebro humano como uma estrutura complexa pré-programada (*prewired*), mas maleável. Ele vale-se da analogia feita pelo neurocientista Gary Marcus (Marcus 2004, p. 12): o cérebro seria como um livro, cuja primeira versão (*first draft*) seria escrita pelos genes no período de desenvolvimento fetal. Nenhum capítulo estaria completo no nascimento, mas também nenhum deles seria composto por páginas em branco nas quais a sociedade pudesse escrever quaisquer palavras (Haidt, 2012, cap. 7).

Outros autores como Mikhail adotam implícita ou explicitamente a ideia, aqui criticada, de que os traços inatos seriam determinados geneticamente:

As this concept will be understood here, nativism is simply the view that early cognitive development is best understood, from a scientific point of view, as a process of biologically determined growth and maturation. As such, nativism assumes that while the acquisition of cognitive systems is triggered and shaped by appropriate experience, and thus depends crucially on cultural inputs, the specific properties of those systems are largely predetermined by the innate structure of the mind. [grifo nosso] (Mikhail, 2011, p. 35)

Como vimos, qualquer desses critérios são insatisfatórios para definir o conceito de inato. Assim, a fim de evitar ambiguidades, evitarei daqui para frente a utilização da palavra ‘inato’ e procurarei mencionar de maneira mais específica, em cada caso, qual propriedade está em discussão. No entanto, algumas vezes, quando estiver reproduzindo o pensamento de outros autores, serei obrigado a utilizar esse termo, quando não estiver muito claro a que propriedade específica eles se referem. O objetivo perseguido nos próximos capítulos é compreender os processos de desenvolvimento e evolução envolvidos na moralidade. A compreensão desses processos envolve discussões a respeito das propriedades normalmente relacionadas ao conceito de inato. Em vez de perguntar se a moralidade é inata, procurarei responder questões dos seguintes tipos: De que maneira os indivíduos aprendem a fazer julgamentos morais? Quais são as características e tendências presentes nos seres humanos que dão suporte a esse aprendizado? Quais são as capacidades intelectuais envolvidas? Que tipo de restrições (*constraints*) atuam sobre a moralidade?

3 Moralidade e cognição

Autores que adotam a abordagem da ética evolucionista para explicar a conformação da psicologia humana por vezes confundem moralidade com altruísmo — psicológico ou biológico²⁰ — ou com comportamento pró-social, ignorando que a existência de comportamentos cooperativos e de punições não depende da existência de juízos morais: ninguém diria que uma abelha ataca quem ameaça a sua colmeia em razão de julgar que tal atitude está moralmente errada e, portanto, merece punição.

Igualar esses conceitos implicaria uma ampliação injustificada do conceito de moralidade que ignoraria o seu caráter normativo. Sensibilidade a situações que causam danos a terceiros, empatia pelos membros do grupo, capacidade para generalização, capacidade para memorizar condutas não-cooperativas, todas essas características podem possibilitar o bom convívio no mundo social sem implicar a existência de moralidade. Podemos imaginar um ser que colabore com os membros de sua espécie, esteja disposto a sacrificar-se conscientemente em benefício de seu grupo, mas seja incapaz de julgar que alguma ação merece punição, ou que ela é desejável e não simplesmente desejada, ou que a autoridade de uma norma independe de convenção (Joyce, 2013, p. 547).

A distinção feita por Darwin (2009) entre “social instincts” e “intellectual powers” é ainda hoje relevante, na medida em que explicita o fato de que é concebível a existência de grupos de cooperadores/altruístas que não possuam faculdade moral (Joyce, 2006). Para que animais se comportem de uma forma ou de outra não é necessário que julguem determinada conduta como boa ou ruim: para que um comportamento seja selecionado basta que os indivíduos o adotem, eles não precisam fazê-lo conscientemente. Isto é, poderia haver condutas que hoje consideramos moralmente louváveis sem que houvesse qualquer ser capaz de fazer julgamentos morais. Poderíamos agir altruisticamente por alguma inclinação sem que acreditássemos dever agir dessa forma. Fazer algo porque gostamos de fazer é diferente de fazer algo por acreditar que temos um dever (Joyce, 2006, p. 50).

Obviamente o fato de moralidade e altruísmo não se confundirem não implica que o surgimento do altruísmo psicológico não tenha sido um precursor essencial do julgamento moral (Joyce, 2013, p. 537). Emoções que favorecem o comportamento social evoluíram antes da cognição moral e tornaram possível a cooperação e o aprendizado cultural, preparando o caminho para que o pensamento normativo fosse possível (Sterelny, 2012).

²⁰ Para a distinção entre altruísmo psicológico e altruísmo biológico ver Sober e Wilson (1999).

Para conseguirmos distinguir claramente esses dois conceitos é indispensável reconhecermos que os julgamentos morais envolvem crenças, em algum sentido relevante. Estaremos então em uma posição melhor para compreender as diversas capacidades intelectuais envolvidas nos julgamentos morais, o que é essencial para o debate em torno do nativismo moral.

Contudo, essa ideia de que julgamentos morais envolvem crenças enfrenta objeções. Essa discussão é um dos pontos centrais do debate metaético entre cognitivistas e não cognitivistas. Essa discussão envolve tanto aspectos conceituais quanto questões empíricas acerca da psicologia humana. No entanto, os argumentos apresentados pelos filósofos da tradição analítica, como regra, enfatizam os aspectos conceituais, e, portanto, não são suficientes para esclarecer esse debate.

Ainda assim, os argumentos empregados pelos cognitivistas são relevantes na medida em que mostram que, uma vez aceitos os pressupostos não cognitivistas, o raciocínio moral não faria qualquer sentido. Se dizer que algo é errado equivale apenas a expressar uma atitude de desaprovação, teríamos dificuldade para explicar a atitude de quem se questiona se as suas próprias atitudes estão moralmente corretas.

Os estudos psicológicos mais recentes acerca dos papéis desempenhados pelas nossas emoções e pela nossa capacidade de raciocínio nos julgamentos morais ajudarão a preencher os espaços deixados pelos argumentos puramente conceituais. Esses estudos indicam que os julgamentos são em larga medida influenciados pelas emoções. No entanto, eles envolvem também processos de raciocínio consciente. A análise desses estudos nos ajudará a compreender as capacidades intelectuais envolvidas nos julgamentos morais.

3.1 Cognitivism e não cognitivism

O não cognitivism apoia-se em duas teses negativas centrais: *não fatualismo semântico* e *não cognitivism psicológico*. A primeira tese nega que sentenças nas quais se atribui predicados morais expressem proposições que possam ser verdadeiras ou falsas em algum sentido substancial, ou seja, elas não possuem condições de verdade. A segunda tese nega que os estados mentais manifestados por meio de discursos morais sejam crenças. Esses estados mentais não seriam ‘cognitivos’, pelos menos não no mesmo sentido que as crenças são. Eles serviriam para expressar atitudes ‘não cognitivas’ similares a desejos e atitudes de aprovação ou desaprovação (van Roojen, 2015; Joyce, 2015). No entanto, as diferentes teorias não-cognitivistas – emotivismo, prescritivismo, expressivismo, *etc.* - divergem em relação ao tipo exato de atitude manifestada pelos julgamentos morais. Por exemplo, conforme A.J. Ayer,

quando dizemos “roubar é errado”, é como se estivéssemos apenas dizendo “roubando dinheiro!!!” com um tom de voz indicando um sentimento de desaprovação (Ayer, 2002, p. 107). Já para Carnap, dizer que matar é errado seria o mesmo que dizer “não mate!” (Carnap, 1979).

Cognitivistas, por sua vez, negam o não cognitivismo. Eles afirmam que as sentenças contendo julgamentos morais estão aptas a serem verdadeiras ou falsas, e que o estado mental de quem emite ou aceita um julgamento moral equivale, como regra, a uma crença. Portanto, sentenças com predicados morais seriam semelhantes a outras sentenças com predicados descritivos (van Roojen, 2015; Mameli, 2013, p. 903-4).

Grande parte do apelo do não cognitivismo decorre de ele representar uma alternativa ao naturalismo moral e a outras formas de realismo. No começo do Século XX, o argumento da questão em aberto de Moore, apesar dos problemas já mencionados no primeiro capítulo, convenceu a muitos de que sentenças contendo expressões morais não eram equivalentes a nenhuma sentença contendo apenas expressões descritivas ou não morais. Não cognitivistas aceitaram esse argumento, mas rejeitaram a conclusão de Moore de que as propriedades morais seriam *sui generis*. Eles optaram por afirmar que os predicados morais simplesmente não se referiam a nenhuma propriedade (van Roojen, 2015). Dessa maneira, o não cognitivismo passou a ser uma opção para aqueles que enxergavam falhas no naturalismo moral mas não concordavam com o intuicionismo proposto por Moore.

Contudo, cognitivistas não precisam ser realistas morais. Como já disse, os realistas morais se caracterizam por defender que propriedades morais estão presentes no mundo para serem descobertas, independentemente das preocupações e práticas dos seres humanos. Antirrealistas, por sua vez, caracterizam-se por negar essa tese. Essa negação pode ocorrer de duas formas: (1) eles podem simplesmente negar que propriedades morais existam; (2) eles podem aceitar que elas existem, mas afirmar que essa existência resulta da atividade mental dos seres humanos (*mind-dependency*). A primeira forma de negação costuma ser adotada por não cognitivistas e por defensores da teoria do erro moral. Já aqueles que adotam a segunda forma de negação podem ser classificados como não objetivistas. Portanto, antirrealistas costumam adotar uma dessas três teses: i – não cognitivismo moral; ii – teoria do erro moral; iii – não objetivismo moral (van Roojen, 2015; Joyce, 2015). Desta forma, o cognitivista ainda pode ser um antirrealista desde que adote a teoria do erro moral ou alguma forma de não objetivismo.

Alguns filósofos contemporâneos que estudam a moralidade a partir de uma perspectiva evolutiva adotam alguma forma de não cognitivismo. Mameli, *e.g.*, defende uma teoria expressivista sobre a moralidade. Segundo ele, a habilidade de manifestar estados mentais a

respeito do que é moralmente exigível é a de adquirir certos tipos de disposições emocionais direcionadas aos violadores de certos tipos de demandas práticas. Ou seja, julgamentos morais, segundo Mameli, são constituídos apenas por disposições para sentir certos tipos de emoção. De acordo com essa visão, se você não está disposto a sentir raiva em relação a um assassino, então você não considera genuinamente que assassinatos são moralmente proibidos (Mameli, 2013, p. 903-7). Prinz adota posição similar à de Mameli. Ele entende que emoções além de *necessárias* são também *suficientes* para avaliarmos algo como bom ou mau; ou seja, valores morais seriam sentimentos e nada mais (Prinz, 2013). Pensar que roubar é errado equivaleria a ter sentimentos negativos frente a esta conduta (Prinz, 2013, p. 103)²¹.

Autores não cognitivistas que adotam posições semelhantes às de Mameli e Prinz têm dificuldade em explicar por que parece tão plausível compreender os julgamentos morais como um modo de expressar crenças e de atribuir propriedades a certas condutas. Eles também têm dificuldade em explicar por que o pensamento sobre questões morais parece se submeter às regras da lógica, ao contrário do que ocorre com as atitudes às quais eles buscam equipará-los – manifestações de emoções, prescrições, etc. (Sayre-McCord, 2014).

O não cognitivista procura explicar os significados de expressões morais quando utilizadas como predicados em sentenças simples (*free standing predicative uses*) e os estados mentais que manifestamos por meio dessas expressões. No entanto, ele tem dificuldade para explicar a utilização de expressões morais em outras sentenças, nas quais os predicados aparecem embutidos em estruturas mais complexas (*embedded sentences*). Dito de outro modo, ele se propõe a explicar os usos de predicados morais quando esses predicados são afirmados diretamente, mas não oferece explicações satisfatórias para os casos em que esses predicados aparecem de outras formas.

Embora possa parecer plausível que a explicação não cognitivista de que o significado de ‘mentir é errado’ é simplesmente uma expressão de aversão à mentira, ela não é uma boa explicação para a atitude manifestada por essa mesma expressão quando ela está embutida em outras sentenças, tais como: ‘é verdade que mentir é errado’; ‘eu me pergunto se mentir é errado’; ‘Paulo acredita que mentir é errado’, etc. É completamente implausível afirmar que quem emprega a expressão ‘mentir é errado’ em sentenças desse tipo está tentando simplesmente expressar uma atitude de aversão. Quem se pergunta se mentir é errado não está manifestando uma emoção negativa em relação ao ato de mentir ou exprimindo um comando como ‘não minta!’ (van Roojen, 2015).

²¹ Prinz tenta explicar o surgimento do julgamento moral a partir da emoção de culpa, mas essa é uma emoção complexa, rica em elementos cognitivos. Ver Joyce (2006, p. 101–4).

O não cognitivista ainda poderia sustentar que os predicados morais possuem um significado quando são afirmados diretamente, e outro quando estão embutidos em outra sentença — afinal as mesmas palavras podem ter usos distintos em ocasiões distintas. No entanto, para ser consistente, o não cognitivista precisará negar a possibilidade de argumentos logicamente válidos serem elaborados a partir de predicados morais. Para que um argumento seja logicamente válido é necessário que as expressões empregadas preservem o mesmo significado ao longo do argumento, caso contrário o argumento será falacioso. Se atribuirmos um significado a um predicado moral quando ele é afirmado e outro quando ele aparece embutido, não conseguiremos sequer elaborar um argumento na forma *modus ponens*. Considere, por exemplo, o seguinte argumento mencionado por Geach:

*If doing a thing is bad, getting your little brother to do it is bad.
Tormenting the cat is bad.
Ergo, getting your little brother to torment the cat is bad (1965, p.463).*

Para que esse argumento seja válido é indispensável que a palavra ‘bad’ tenha o mesmo significado nas quatro vezes em que é mencionada. Ela não pode ser descritiva em um dos usos, manifestar uma emoção ou expressar um imperativo em outro e assim por diante. Ora, está claro que a palavra ‘bad’ presente na premissa maior não está ali para expressar uma atitude de condenação ou algo semelhante: ninguém diria que na primeira parte dessa premissa se está simplesmente condenando o ato de ‘doing a thing.’ Desta forma, uma vez que o significado de ‘bad’ deve permanecer constante, também não se poderia dizer que a premissa menor expressa simplesmente uma atitude de condenação ao ato de torturar um gato (Geach, 1965, p.464).

De modo mais geral, para que qualquer argumento da forma ‘se *p*, então *q*; *p*; logo *q*’ seja válido, é indispensável que *p* tenha o mesmo significado quando aparece sozinho e quando aparece embutido na premissa ‘se *p*, então *q*.’ No entanto, o não cognitivista que diz que os predicados morais significam uma coisa quando afirmados isoladamente e outra quando estão embutidos, diz justamente que *p* (*e.g.*, mentir é errado) tem um significado quando está isolada e outro quando está embutida (*e.g.*, se mentir é errado, então ...).

Não há dúvidas, como veremos na seção seguinte, de que os julgamentos morais são influenciados pelas emoções e servem muitas vezes para expressar atitudes, mas isso não nos permite concluir que esses julgamentos não refletem crenças e não podem ser avaliados como falsos ou verdadeiros (*truth-evaluable judgements*). É difícil explicar até mesmo certas emoções sem recorrermos a esse tipo de julgamento no qual estão presentes condições de verdade. A emoção de culpa, *e.g.*, não envolve apenas um sentimento (*feeling*), mas também certos pensamentos como “eu transgredi uma norma” ou “eu não cumpri com as minhas obrigações” (Joyce, 2016, p. 1-2).

Inclusive valorações feitas a partir de uma perspectiva puramente egocêntrica costumam ser parcialmente descritivas. Obviamente elas não são puramente descritivas, pois um elemento essencial desse tipo de valoração é a aprovação, por parte de quem julga, de certos requisitos — estejam eles explícitos ou não — que o objeto avaliado supostamente possui. Mas ela é parcialmente descritiva, pois afirma que o objeto avaliado possui as características necessárias para satisfazer esses requisitos, sejam eles quais forem. Quando alguém que acredita na existência de valores morais objetivos julga moralmente, o caráter descritivo do julgamento se torna ainda mais claro. Quando essa pessoa diz que alguma conduta é boa, ela pretende afirmar algo sobre o mundo, *i.e.*, o suposto fato de que essa conduta satisfaz certos requisitos intrínsecos (Mackie, 1990, p. 61-3). Se, *e.g.*, um utilitarista diz que mentir é errado, posso inferir que ele acredita que não mentir ajuda a maximizar a utilidade.

Críticas como essa a respeito do significado de expressões morais quando embutidas levaram a uma reformulação das posições não cognitivistas. O não cognitivismo surgiu como uma alternativa em relação às, então predominantes, teorias cognitivistas e realistas. À proporção que seus defensores tentaram explicar determinadas características do discurso moral que serviam de base para os seus críticos, as teorias não cognitivistas tornaram-se mais sutis e menos contrastantes com as posições contrárias. As suas duas teses centrais (*não fatalismo semântico* e *não cognitivismo psicológico*) tornaram-se mais moderadas. A tese de que os julgamentos morais não têm significado descritivo passou a ser a de que esse significado existe, mas é secundário. A tese de que esses julgamentos não podem ser verdadeiros ou falsos tornou-se a de que eles podem ser verdadeiros ou falsos apenas em um sentido mínimo²². Naturalmente, essa transição não foi aceita por todos os não cognitivistas, mas ela é uma tendência bastante nítida nas teorias mais recentes (van Roojen, 2015).

As teorias expressivistas híbridas, *e.g.*, combinam ideias expressivistas e cognitivistas. Desta forma, as sentenças morais seriam instrumentos convencionais utilizados para expressar atitudes de aprovação ou desaprovação, mas ao mesmo tempo elas serviriam para atribuir propriedades; e as atitudes morais seriam parcialmente cognitivas. Qual a propriedade exata atribuída pelo julgamento moral dependeria das atitudes morais de quem o profere. No entanto, quem atribui predicados por meio de julgamentos morais não pretende apenas descrever o mundo, mas expressar o seu posicionamento perante aquilo que é descrito (van Roojen, 2015).

Como resultado da crescente semelhança entre teorias cognitivistas e não cognitivistas, os argumentos para sustentar uma ou outra posição foram se tornando cada vez mais intrincados. Chegamos a um ponto em que é difícil dizer se as teorias elaboradas para lidar com

²² A respeito da concepção minimalista de verdade, ver Ritchie (2008, p. 175).

as críticas feitas pelos cognitivistas ainda merecem ser rotuladas como não cognitivistas. Filósofos importantes sugerem que a distinção entre cognitivismo e não cognitivismo não tem mais utilidade (Gibbard, 2003, p. 183–84; Korsgaard, 2008, p. 309). Talvez essa distinção tenha colapsado à medida que as teorias não cognitivistas se modificaram a fim de responder as críticas feitas pelos cognitivistas (van Roojen, 2015).

As teorias não cognitivistas serviram para colocar em evidência que os julgamentos morais desempenham outros papéis além daqueles que os cognitivistas lhes costumavam atribuir. Hare, *e.g.*, estava certo ao afirmar que expressões valorativas são utilizadas com a função de recomendar algo (Hare, 1991, p. 91). No entanto, como os defensores de certas teorias híbridas e de algumas teorias não cognitivistas mais sofisticadas perceberam, as expressões valorativas podem também possuir um significado descritivo e expressar crenças.

Não pretendo aqui ser o árbitro dessa discussão. A forma como procurei caracterizar a moralidade poderia ser aceita por cognitivistas de diversos matizes e também por aqueles que adotam posições não cognitivistas mais amenas. Para os assuntos que serão abordados nos próximos capítulos, é desejável apenas que fique claro que os julgamentos morais envolvem crenças e possuem um sentido descritivo, ainda que eventualmente secundário.

3.2 Contribuições da Psicologia Moral

A produção de julgamentos morais envolve o funcionamento de diferentes sistemas cognitivos e ocorre em diferentes etapas, algumas das quais são inacessíveis à introspecção. Dentre os pesquisadores dedicados ao estudo da moralidade sob uma perspectiva evolutiva há praticamente um consenso sobre esse ponto. No entanto, não há acordo sobre a importância relativa de cada uma dessas etapas.

Haidt e Bjorklund acreditam que o julgamento moral é o produto de intuições rápidas e automáticas, as quais são seguidas por um lento processo de raciocínio moral, cuja função principal é encontrar razões para justificar essas intuições iniciais. O raciocínio do próprio indivíduo raramente teria algum valor causal direto sobre o julgamento moral por ele realizado, exceto em subculturas muito específicas como a dos filósofos (2008, p. 193). Segundo, esses autores, nós frequentemente apresentamos argumentos como mero recurso retórico para convencer a nós mesmos de que estamos certos ou para conseguir aliados.

As justificações morais teriam, entretanto, efeito causal quando empregadas em discussões morais travadas com outras pessoas. Há uma série de pesquisas na área de psicologia social que demonstram a tendência de sintonizarmos as nossas atitudes àquelas das pessoas com as quais interagimos (*e.g.*, Pennington & Schlenker, 1999; Sinclair *et al.*, 2005). Nós sentimos

uma necessidade de pertencer a um determinado grupo e os nossos julgamentos morais são moldados por aquilo que os outros membros do grupo acreditam. Além disso, nós somos péssimos em procurar evidências contrárias às nossas crenças, mas somos muito bons em encontrar falhas nas crenças das pessoas ao nosso redor. Assim, frequentemente só conseguimos enxergar os argumentos contrários às nossas crenças quando outros membros do nosso grupo os explicitam. Por esses motivos, o julgamento moral precisa ser compreendido também como um processo social, não apenas como um ato privado de cognição (Haidt & Bjorklund, 2008, p. 181; Haidt, 2012).

O cérebro avalia e reage incessantemente. Há tecidos nervosos responsáveis por integrar informações oriundas dos ambientes interno e externo e responder a uma questão básica: aproximar-se ou evitar (*approach-avoid*)? No caso humano, essa questão traduz-se na dicotomia bom e mau. Esse tipo de avaliação (bom-mau) é a dimensão mais básica de todos os julgamentos. A mente humana está o tempo todo avaliando e julgando tudo o que ela vê e ouve em termos de bom ou mau, mesmo quando se trata da face de uma pessoa ou de palavras em uma língua estrangeira (Haidt & Bjorklund, 2008, p. 186-7).

Aquilo que Haidt chama de intuição moral se refere ao surgimento repentino na consciência de um sentimento valorativo (bom-mau) em relação ao caráter ou às ações de outrem, sem que o sujeito tenha conscientemente seguido etapas de pesquisa, ponderado evidências ou inferido conclusões. A intuição moral é um tipo especial de processo automático, no qual sempre está envolvido, ao menos, algum traço de sentimento valorativo. Na maioria dos casos, essa intuição (*flash of feeling*) conduzirá diretamente à condenação ou à aprovação da conduta em questão (Sinnott-Armstrong, Young & Cushman, 2010, p. 246). Haidt denomina julgamento moral essa experiência consciente de censura ou elogio, que inclui a crença a respeito da retidão do ato. Ou seja, como regra, o julgamento moral apenas confirmaria a intuição moral (Haidt & Kesebir, 2010, p. 802; Haidt & Bjorklund, 2008, p. 188).

Por outro lado, o raciocínio moral é uma atividade mental consciente, caracterizada pela transformação de informações recebidas sobre pessoas e situações com a finalidade de alcançar um julgamento moral. Qualificar o raciocínio moral como um processo consciente significa afirmar que ele é intencional, controlável e exige esforço de quem o realiza (Haidt, 2001, p. 818).

Haveria, então, pelo menos dois tipos distintos de processos mentais envolvidos nessas valorações: um deles seria automático e o outro controlado. A maior parte da cognição humana pertence ao primeiro tipo e é semelhante à cognição presente em outros animais. Trata-se de um sistema muito antigo, surgido há milhões de anos, cujo funcionamento permite a solução de

inúmeros problemas complexos de forma inconsciente. Ele possibilita, por exemplo, o reconhecimento de pessoas e objetos instantaneamente. Por outro lado, há um tipo de cognição que é único ao ser humano. Ele opera por meio de palavras e, portanto, não pode ser mais antigo do que a linguagem. Ao contrário do processo anterior, a habilidade para raciocinar por meio da linguagem é um processo lento e exige esforço. Haidt emprega a metáfora do elefante que é conduzido por um homem para explicar como esses diferentes tipos de cognição interagem (Haidt, 2013)²³:

Automatic cognition -the elephant- is sometimes called “hot” cognition because it has the power to motivate us to action. The neurological systems for judgement are connected to the neurological systems for behavior. Controlled cognition, in contrast, is sometimes called “cool” cognition. It is not connected to behavioral centers in the brain. So please imagine the human mind as a small and somewhat ineffectual rider perched on the back of a large, powerful, and rather smart elephant. The rider can try to steer the elephant, and if the elephant has no particular desire to go one way or the other, it may listen to the rider. But if it has its own desires, it’s going to do what it wants to do (Haidt, 2013).

Há uma série de pesquisas consistentes com a ideia de que muitas vezes avaliamos instantaneamente e depois raciocinamos para justificar a avaliação feita, e de que este processo consciente de raciocínio poucas vezes inibe ou altera a decisão tomada (Zajonc, 1980; Monahan *et al.*, 2000; Luo *et al.*, 2006). Dentre os estudos consistentes com esta ideia, há pesquisas empíricas realizadas através de entrevistas sobre julgamentos morais (Haidt *et al.*, 1993); de questionários respondidos pela Internet sobre os *trolley problems* (Hauser *et al.*, 2008); de manipulação de julgamentos morais por meio da indução de emoções com a utilização de hipnose e de alterações ambientais (Wheatley & Haidt, 2005; Schnall *et al.*, 2008); de imagens por ressonância magnética funcional (Greene & Haidt, 2002).

Talvez o exemplo mais claro de como a consciência fornece racionalizações *post hoc* seja fornecido pelos estudos de Gazzaniga (1985) em pacientes que tiveram o corpo caloso responsável pela conexão dos dois hemisférios cerebrais seccionados. Quando um paciente realizava uma ação em função de um estímulo sobre o hemisfério cerebral direito, *e.g.*, levantar e sair andando, o hemisfério esquerdo, responsável pela linguagem, não respondia algo como ‘não sei por que estou fazendo isso’, mas inventava uma razão qualquer, como ‘estou indo buscar um refrigerante.’ Segundo Gazzaniga, nosso raciocínio verbal consciente não comanda as nossas ações; ele funciona como uma secretária de imprensa encarregada de inventar explicações convincentes para qualquer coisa que a pessoa faça. Como não temos acesso ao sistema de cognição rápida que desencadeia nossos julgamentos e condutas, só nos restaria

²³ Greene faz analogia similar, ao comparar os diferentes tipos de cognição com os modos automático e manual das câmeras fotográficas (2013, p. 133–4).

empregar o sistema de cognição lenta para adivinhar explicações para os seus resultados (Gazzaniga, 1985; Haidt & Bjorklund, 2008; Wahl, 2014, p. 69-71).

No entanto, como veremos adiante, há casos em que os processos conscientes de raciocínio parecem sim desempenhar algum papel causal sobre a produção dos nossos julgamentos morais.

3.2.1 Emoção e julgamento moral

Certos estudos pretendem analisar o desenvolvimento e a evolução da moralidade, mas se limitam a tratar de sentimentos morais e ignoram diversos aspectos cognitivos. Outros restringem a moralidade à capacidade para formular julgamentos morais com base em conceitos abstratos. Acredito que nenhum dos dois enfoques seja adequado, pois a moralidade envolve diferentes aspectos emocionais e cognitivos.

Nas últimas décadas, vem ganhando força a tradição sentimentalista (*moral sense*), iniciada por filósofos ingleses e escoceses, como o Conde de Shaftesbury, Francis Hutcheson, Adam Smith e David Hume (Hauser, 2007, p. 23). Dentre os filósofos que defendem essa tradição, destacamos Kim Sterelny:

I am allied with a Humean revival: we are perceptually tuned to the emotions and emotional responses of others, and to our own emotional responses to those others. We are typically aware of, and emotionally responsive to, others' distress. We do not just respond emotionally; we notice our own emotions by some internal analogue of perception. These features of emotion are the launching pad for the Humean revival in psychology and philosophy known as sentimentalism. (...) Moral norms grow from our dispositions to respond emotionally in characteristic ways to stereotypical stimuli: for example, with disgust to body products and rotting food, or with distress and discomfort to the suffering of others (or at least of those in our group) (Sterelny, 2012).

Haidt também se filia a essa tradição. Ele acredita que as nossas motivações e crenças morais surgem a partir de sentimentos característicos da condição humana, os quais são responsáveis pelas nossas intuições a respeito do certo e do errado (Haidt & Bjorklund, 2008, p. 185). A posição de Haidt situa-se em um dos polos de uma controvérsia presente na área da psicologia, desde o final dos anos 80, entre teorias racionalistas, baseadas nos trabalhos de Piaget e Kohlberg, e teorias intuicionistas (Haidt & Bjorklund, 2008, p. 185; Hauser *et al.*, 2008, p. 114). No âmbito da psicologia moral, predomina atualmente o entendimento segundo o qual há uma predominância da intuição, ainda que tanto o raciocínio quanto a intuição sejam reconhecidos como relevantes (Haidt & Kesebir, 2010, p. 807). Há, no entanto, importantes pesquisadores contrários a esta mudança em direção ao intuicionismo (Killen & Smetana, 2005).

O filósofo Jesse Prinz (2013) relaciona vários estudos empíricos recentes na área da ciência cognitiva que reforçam a tradição sentimentalista. Esses estudos indicam que a emoção ocupa uma posição central nos julgamentos morais. Estudos com neuroimagem mostram que centros cerebrais responsáveis pelas emoções são ativados quando resolvemos dilemas morais (Greene *et al.*, 2001), lemos descrições de violações morais (Moll *et al.*, 2002a), vemos imagens com significados morais (Moll *et al.*, 2002b), ou confrontamos jogadores de moralidade questionável em jogos econômicos (Sanfey *et al.*, 2003).

Apesar dos resultados dessas experiências, os racionalistas ainda poderiam estar certos. Essas pesquisas demonstram apenas que as emoções são ativadas quando fazemos julgamentos morais; elas não demonstram que as emoções exercem um papel causal sobre a produção desses julgamentos. De acordo com a visão racionalista, julgamos moralmente as condutas que observamos sem a necessidade de consultarmos as nossas emoções. Deste modo, os racionalistas poderiam dizer que a presença de emoções é apenas um efeito da produção de julgamentos morais. Portanto, a simples comprovação da presença de emoções não é suficiente para optarmos pela posição sentimentalista (Prinz, 2012).

No entanto, há também diversos estudos psicológicos que demonstram que as emoções influenciam os julgamentos morais. Estudos comportamentais demonstram como as pessoas fazem julgamentos morais mais severos quando estão em ambientes sujos, fétidos e desorganizados, ou quando um sentimento de raiva é induzido por meio de filmes ou de recordações autobiográficas (Schnall *et al.* 2008; Lerner *et al.* 1998). Em outra experiência, Kendall Eskine demonstrou a influência de ingerir bebidas amargas nos julgamentos morais (Eskine *et al.*, 2011). Além da emoção de nojo, que faz as pessoas serem menos dispostas a perdoar certas condutas, a raiva induzida também aumenta a vontade de punir. Estudos também demonstram como a indução do sentimento de alegria por meio de filmes de comédia pode induzir as pessoas a serem mais utilitaristas, *e.g.*, favorecendo o sacrifício de uma pessoa inocente para salvar outras cinco em perigo (Valdesolo & DeSteno, 2006).

As emoções que sentimos funcionam muitas vezes como um instrumento heurístico para identificar se algo deve ser condenado moralmente – *affect heuristic* (Kahneman & Frederick, 2005, p. 271-283; Slovic *et al.*, 2007). A aplicação da *affect heuristic* segue o seguinte padrão: se pensar sobre a prática de determinado ato provoca um mal-estar, então julgue que esse ato é moralmente errado. Frequentemente substituímos, ainda que de forma inconsciente, a forma como nos sentimos perante um ato pelo valor moral desse ato, *i.e.*, julgamos que o ato é moralmente errado com base na forma como nos sentimos (Sinnott-Armstrong, Young & Cushman, 2010, p. 259-61).

Nesse momento, racionalistas poderiam conceder que as emoções eventualmente influenciam os julgamentos morais, mas defender que essa influência é apenas uma interferência, não algo intrínseco aos julgamentos morais. Contudo, há estudos que indicam que a mera presença de certas emoções pode desencadear julgamentos morais em situações nas quais não há qualquer princípio racional disponível para sustentá-los. Wheatley e Haidt, *e.g.*, conduziram um estudo no qual as pessoas foram induzidas, por meio de hipnose, a sentir nojo quando ouvissem uma determinada palavra ('often' ou 'take'). Após a sessão de hipnose, os pesquisadores apresentaram diferentes estórias aos participantes e pediram para que eles julgassem as condutas dos personagens. Os julgamentos relacionados às estórias nas quais a palavra-chave estava presente foram muito mais severos. No entanto, o aspecto mais surpreendente foi o fato de os participantes condenarem a conduta de um personagem que não havia cometido qualquer transgressão. A estória era a seguinte: "Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion" (Wheatley & Haidt, 2005). Haidt descreve a surpresa causada pelo resultado da pesquisa:

We predicted that subjects who felt a flash of disgust while reading this story would have to overrule their gut feelings. To condemn Dan would be bizarre. Most of our subjects did indeed say that Dan's actions were fine. But a third of the subjects who had found their code word in the story still followed their gut feelings and condemned Dan. They said that what he did was wrong, sometimes very wrong. Fortunately, we had asked everyone to write a sentence or two explaining their judgements, and we found gems such as "Dan is a popularity-seeking snob" and "I don't know, it just seems like he's up to something." These subjects made up absurd reasons to justify judgements that they had made on the basis of gut feelings—feelings Thalia had implanted with hypnosis (Haidt, 2012).

As pesquisas de Haidt demonstram o fenômeno chamado *moral dumbfounding*, *i.e.*, essas pesquisas apresentam inúmeros casos — praticar incesto utilizando dois métodos contraceptivos, masturbar-se com galinha morta, *etc.* — nos quais as pessoas condenam moralmente uma conduta, mas não conseguem encontrar qualquer razão para justificar o julgamento, apenas o nojo despertado pela conduta (Haidt, 2001). A única justificativa é o sentimento de que algo está errado.

Em seu livro *Descartes' Error*, Damásio (2008) demonstrou que as áreas do córtex pré-frontal responsáveis por integrar as emoções ao processo de tomada de decisão são extremamente relevantes para os julgamentos e comportamentos morais. Damásio analisou casos de pacientes com lesões no córtex pré-frontal ventromedial (VMPC). Dentre estes, destaca-se o famoso caso de Phineas Gage, operário norte-americano que, em um acidente

envolvendo a utilização de explosivos ocorrido em 1848, teve a cabeça transpassada por uma barra de ferro. Gage não ficou com sequelas físicas aparentes, com exceção da visão do olho esquerdo, atingido pela barra, e sua capacidade para o raciocínio lógico também não sofreu prejuízos relevantes. Contudo, houve mudanças significativas em sua conduta, e ele passou a ter dificuldades para ajustar-se às normas morais da sociedade em que vivia. Para descrever a mudança de personalidade observada em decorrência de danos cerebrais desse tipo, tem sido utilizado o termo ‘sociopatia adquirida’ (Del-Ben, 2005).

Pacientes com lesões semelhantes àquela de Gage apresentam redução de empatia e de afeto, mas mantêm outras capacidades intelectuais praticamente inalteradas. Por esse motivo, eles fornecem uma ótima oportunidade para estudar o papel das emoções nos julgamentos morais. Aparentemente, eles perdem a capacidade para integrar respostas afetivas com a cognição envolvida nos julgamentos morais. Estudos feitos com pacientes que sofreram danos no VMPC, tanto na infância quanto durante a vida adulta, indicam que a adesão às normas sociais depende de controle emocional. Este tipo de controle, por sua vez, depende da existência de conexões intactas entre os lobos frontais e o sistema límbico – em especial a amígdala, responsável pela expressão das emoções (Hauser, 2007, p. 227; Cushman, Young & Greene, 2010, p. 53-4).

Pacientes com as lesões mencionadas são capazes de reconhecer quais são as normas vigentes, mas, em razão da ausência de certas emoções, eles têm dificuldades para decidir qual conduta adotar e, quando decidem, muitas vezes o fazem em desacordo com os padrões considerados normais. Em certos cenários, eles tendem a optar por soluções muito mais utilitaristas do que aquelas adotadas por sujeitos saudáveis. Essa diferença é marcante nos cenários classificados como pessoais em oposição aos cenários impessoais. Nesses cenários, pacientes com lesão no VMPC tendem a aprovar a prática de condutas lesivas que usualmente despertam forte aversão emocional se essas condutas beneficiam um número maior de pessoas. Em dilemas semelhantes aos *trolley problems*, e.g., eles tendem a sacrificar a vida de uma pessoa para salvar outras cinco ainda que essa conduta envolva a prática de comportamentos considerados aversivos emocionalmente, como derrubar alguém de uma ponte (Koenigs *et al*, 2007; Ciaramelli *et al*, 2007). Ao que tudo indica, a produção de julgamentos deontológicos deve-se em grande medida a respostas emocionais intuitivas que dependem do VMPC (Cushman, Young & Greene, 2010, p. 53-4).

Estudos como esses indicam a natureza causal das emoções em relação aos julgamentos morais (Sinnott-Armstrong, Young & Cushman, 2010, p. 265). Contudo, embora as emoções

sejam muito relevantes, talvez elas não sejam necessárias; eventualmente, em certas circunstâncias, sejamos capazes de julgar moralmente apenas a partir de princípios racionais.

Hauser, por exemplo, acredita que déficits nas emoções afetam mais diretamente a performance moral e não tanto a capacidade para julgar moralmente. Isto é, indivíduos com este tipo de déficit seriam capazes, em diversas situações, de julgar qual a conduta correta a ser adotada. No entanto, em razão da ausência das emoções relevantes, eles teriam dificuldades em aderir a esse julgamento e guiar as suas condutas por ele (Hauser, 2007, p. 229).

Embora pacientes com lesões no lobo frontal julguem de maneira mais utilitarista os *trolley problems*, em outros casos os seus julgamentos são semelhantes àqueles de outros sujeitos (Koenigs *et al.*, 2007):

When they read the philosopher Peter Unger's altruism cases—driving by the injured child on the side of the road versus throwing the UNICEF card away—they were like normal subjects: You can't drive by the injured child on the road but you can toss the UNICEF card. They were also like normal subjects for the CEO cases described in chapter 1: They attribute intent and blame to a CEO who supports a large moneymaking policy that hurts the environment, but they do not make this attribution when the policy ends up helping the environment. Last, these patients were more likely than normal subjects to show sympathy and generosity in cases involving the homeless. For example, they were more likely to say that a person walking home on a blistery cold night, one block from home, should give up his winter coat to a homeless man or collection box, an act that will save the homeless person from freezing to death overnight (Hauser, 2007, p. 231-2).

Essa observação sugere que o quadro sobre a influência dessas lesões sobre os julgamentos é mais complexo.

No entanto, estudos com crianças que sofreram lesões no córtex orbitofrontal indicam que as emoções podem ser também indispensáveis para o aprendizado moral (Haidt & Bjorklund, 2008, p. 199). Crianças com esse tipo de lesão não desenvolvem um *moral sense*, embora desenvolvam as outras funções cognitivas consideradas avançadas – linguagem, planejamento, raciocínio (Anderson *et al.*, 1999). Se os racionalistas estivessem certos, pacientes com lesões no VMPC seriam os melhores candidatos para o papel de agentes morais ideais, em virtude de estarem protegidos das influências emocionais. Neles, a comunicação entre a ‘alma racional’ e as ‘paixões do corpo’ foi destruída. Logo, a razão poderia operar sem ser perturbada por conselhos incautos. No entanto, o que se observa é que na ausência da emoção a razão não sabe para onde ir (Prinz, 2012).

O que foi afirmado em relação à sociopatia adquirida aplica-se também a outros casos de transtorno de personalidade antissocial. Pesquisas feitas sobre sociopatia indicam déficits específicos em emoções relacionadas à moralidade. Sociopatas são desprovidos das reações emocionais que fazem com que as pessoas se importem com essas normas da sociedade e as

internalizem (Blair, 2007; Kiehl, 2006). Tanto aqueles que estudam os comportamentos como aqueles que estudam os cérebros dos sociopatas concordam que a deficiência central, da qual podem ser derivados os outros sintomas, é a incapacidade para sentir emoções, como empatia, vergonha, culpa, que fazem as pessoas se importarem com os destinos umas das outras (Haidt & Kesebir, 2010, p. 804).

Apesar disso, Joyce defende que aparentemente emoções não são necessárias para a realização de julgamentos morais. Ele menciona julgamentos morais feitos em situações de fadiga emocional, como a de alguém submetido a uma sequência de dilemas morais similares (*e.g.*, em um experimento psicológico envolvendo dezenas de cenários semelhantes), a tal ponto que nenhuma emoção seja despertada. Além disso, Joyce salienta que aqueles que afirmam a necessidade de um impulso emocional para ocorrer um julgamento moral focam em casos extremos (pedofilia, trens descarregando passageiros em Auschwitz, *etc.*). Situações comuns do cotidiano — como o julgamento de que o medalhista de ouro merecia vencer, ou de que a propriedade de um sapato garante certos direitos sobre esse acessório — não seriam levadas em consideração pelos sentimentalistas (Joyce, 2014, p. 132-3).

No entanto, valorações parecem ser possíveis somente a partir de sentimentos; sem eles nós podemos categorizar coisas, mas não podemos avaliá-las como boas ou ruins. Quando experimentamos um vinho, *e.g.*, dizemos que ele é bom se bebê-lo nos causa uma emoção positiva. Caso não tenhamos qualquer prazer em bebê-lo, podemos ainda dizer que é um bom vinho em razão de ter sido recomendado por algum *sommelier*. Nessa hipótese, a nossa avaliação é parasitária do gosto do *sommelier* (Prinz, 2013, p. 100). Se buscarmos a origem da avaliação acabaremos encontrando um sentimento.

Quando decidimos, nós analisamos uma série de possíveis consequências envolvidas na decisão e projetamos nossas emoções. A opção escolhida é o fruto de uma combinação de processos conscientes e inconscientes, afetivos e racionais (Haidt & Bjorklund, 2008, p. 195). A simples ‘razão’ não leva a qualquer resultado. Ela só nos diz como agir quando já sabemos o objetivo que queremos atingir (Prinz, 2012).

Portanto, ainda que estejamos dispostos a conceder que o tipo de julgamento feito por sociopatas pode ser chamado de ‘julgamento moral’, é importante destacar que em um mundo composto apenas por sociopatas julgamentos morais não existiriam, pois os seus julgamentos são parasitários em relação a outros julgamentos.

A raiva nos motiva a retaliar aqueles que praticam condutas que consideramos imorais, e a existência de retaliação promove a observância das normas morais vigentes. Nos casos em que não há possibilidade de retaliação, a antecipação da emoção de culpa muitas vezes nos demove

da intenção de transgredi-las. Sem a existência dessas emoções, a observância das normas morais cairia drasticamente. Portanto, ainda que as emoções não sejam indispensáveis para a produção de um julgamento moral, como defende Joyce, elas parecem ser indispensáveis para a continuidade da moralidade no âmbito do grupo, em razão de seu papel motivacional (Prinz & Nichols, 2010, p. 141).

De qualquer modo, com exceção de casos muito específicos, como o dos sociopatas, todos os estudos de neuroimagem mostram que as áreas do cérebro envolvidas no processamento de emoções estão ativas quando fazemos julgamentos morais (Hauser, 2007, p. 223). Como já disse, há inúmeros estudos psicológicos que indicam que emoções não apenas estão presentes quando tentamos solucionar dilemas morais, mas também desempenham um papel causal nas nossas decisões²⁴; e há ainda estudos que indicam que a mera presença de certas emoções pode desencadear julgamentos morais, mesmo quando as pessoas não conseguem fornecer qualquer razão para justificá-los²⁵. Portanto, como regra, há um vínculo muito forte entre as emoções e os julgamentos morais.

3.2.2 Aspectos cognitivos da moralidade

Na seção anterior destaquei a importância da presença de emoções para os julgamentos morais. Cabe agora salientar que não parece adequado igualar a moralidade a emoções, como fazem autores como Prinz. A posição defendida por ele é típica dos antinativistas, que tendem a compreender o julgamento moral apenas em termos de traços emocionais os quais eles acreditam desempenhar funções mais gerais.

No entanto, essa descrição dos julgamentos morais ignora certos aspectos cognitivos (Joyce, 2013, p. 548). Julgamentos morais exigem crenças e ideias como merecimento, capacidade de agir, desaprovação e transgressão, as quais não podem ser explicadas apenas com referências a mecanismos puramente conativos (Joyce, 2013, p. 550; Joyce, 2014, p. 135-137). A internalização de uma norma exige a crença de que uma punição é merecida em caso de transgressão. Somente tendo em vista a existência de elementos cognitivos podemos explicar por que normalmente não fazemos julgamentos morais a respeito de animais. Em razão de não possuírem um tipo específico de capacidade para agir que consideramos essencial para a valoração moral, não condenamos moralmente as ações deles.

²⁴ Por exemplo: Schnall *et al.* 2008; Lerner *et al.* 1998; Eskine *et al.*, 2011; Valdesolo & DeSteno, 2006; Kahneman & Frederick, 2005, p. 271-283; Slovic *et al.*, 2007; Koenigs *et al.*, 2007; Ciaramelli *et al.*, 2007; Anderson *et al.*, 1999.

²⁵ Por exemplo: Wheatley & Haidt, 2005; Haidt, 2001; Haidt, 2012.

Além disso, o raciocínio não serve apenas para racionalizar decisões previamente adotadas. Ainda que, como sustenta Haidt, o raciocínio consciente apenas em poucos casos tenha um papel causal na produção de julgamentos morais, ele pode ser bastante relevante. Uma única decisão tomada por meio desse tipo de raciocínio pode exercer influência duradoura quando altera as motivações que orientam a conduta do indivíduo. Muitos indivíduos, *e.g.*, por meio de raciocínio consciente, chegam à conclusão de que é imoral comer carne e isso acaba produzindo alterações no comportamento que se ramificam ao longo do tempo (Pizarro & Bloom, 2003).

Alterações como essa, principalmente por meio de processos de interação social, influenciarão as emoções e as intuições morais despertadas no futuro. Exemplos como o do vegetarianismo moral demonstram como alterações nos julgamentos morais podem ser provocadas por processos de deliberação e de reflexão (Sterelny, 2012). O nojo e a repulsa que vegetarianos sentem por carne é uma consequência, não uma causa, de suas convicções morais (Fessler *et al.*, 2003). De forma semelhante, pessoas que consideram a homossexualidade repugnante podem tornar-se tolerantes em razão de processos de reflexão e modificar as suas intuições e sentimentos iniciais. Sterelny menciona vários exemplos da história recente que demonstram como as intuições dos indivíduos de uma sociedade podem ter mudanças drásticas em poucos anos. Além da postura em relação à homossexualidade, as respostas à crueldade com animais, aos métodos de educação das crianças, ao papel das mulheres, à obesidade e à depressão sofreram importantes alterações que se refletem nas nossas reações viscerais, não apenas no plano da consciência (Sterelny, 2010, p. 286-7). Casos como esses demonstram a instabilidade e o dinamismo de nossas intuições morais. A mudança de convicção atrelada a uma adequação das práticas cotidianas à nova convicção acabam por alterar os tipos de objeto que provocam repugnância no indivíduo (Sterelny, 2010, p. 282-4 e 293-4; Skorupski, 1993, p. 132-4).

Nesse sentido, até mesmo Prinz reconhece que o raciocínio consciente pode desempenhar algum papel relevante na produção de julgamentos morais. Prinz acredita que temos valores básicos – não matar, não ser cruel, ser justo, *etc.* – que emergem diretamente de nossas emoções; e valores derivados desses valores básicos. Não comer carne, *e.g.*, seria um valor derivado. Aqueles que acreditam que comer carne é errado argumentam que essa conduta é uma forma de crueldade. O papel do raciocínio seria identificar qual valor básico se aplica ao caso em questão (Prinz, 2012).

Há, além disso, uma literatura substancial sobre atitudes sociais implícitas — em especial sobre atitudes raciais — que corrobora a tese de que processos de raciocínio consciente

podem prevalecer sobre processos intuitivos. Já é algo bastante consolidado nessa literatura que uma pessoa pode não manifestar qualquer atitude preconceituosa em testes que envolvem medições explícitas, como questionários, por exemplo, mas apresentar vieses racistas em testes de associação implícita. Na versão mais comum desse teste, figuras de rostos humanos, negros ou brancos, são associadas com palavras consideradas positivas ou negativas. Os participantes tipicamente associam mais facilmente pessoas brancas a palavras positivas e pessoas negras a palavras negativas. Quando elas precisam fazer a associação contrária, *i.e.*, associar pessoas negras a palavras positivas ou pessoas brancas a palavras negativas, costumam demorar mais tempo (*e.g.*, Greenwald *et al.*, 1998; Mallon & Nichols, 2010, p. 300-2).

Essa literatura indica que processos controlados conscientemente podem exercer uma importante influência sobre as respostas verbais e sobre o comportamento. Quando esses processos são sobrecarregados, por meio de tarefas que requerem um acréscimo da atividade cognitiva, ou então quando eles são enfraquecidos, *e.g.*, por meio do consumo de álcool, as respostas verbais e o comportamento passam a se alinhar mais diretamente com as atitudes implícitas.

Richeson e Shelton (2003), *e.g.*, conduziram um experimento destinado a examinar a influência de interações inter-raciais no funcionamento cognitivo. Nessa experiência, os participantes brancos tinham que manter contato por certo tempo com outro indivíduo, branco ou negro, e, então, realizar um teste que exigia atenção e esforço cognitivo (teste de Stroop). Antes da experiência, as atitudes raciais automáticas dos participantes foram medidas por testes de associação implícita. Aqueles que apresentaram vieses racistas mais acentuados nessa etapa prévia obtiveram piores resultados no teste cognitivo. No entanto, esse efeito só foi observado naqueles que haviam interagido com um indivíduo negro. Isto é, aqueles participantes que apresentavam vieses racistas acentuados, mas que interagiram somente com o indivíduo branco obtiveram resultados semelhantes àqueles que não manifestavam esses vieses racistas de modo acentuado. De acordo com a interpretação dos autores do experimento, a interação com o indivíduo negro exigia dos participantes mais preconceituosos uma grande utilização de recursos cognitivos para suprimir as respostas preconceituosas automáticas. Essa alta demanda cognitiva acabava exaurindo os recursos necessários para completar o teste e, assim, prejudicava o desempenho nos testes subsequentes²⁶.

Além disso, uma experiência liderada por Bartholow (Bartholow, Dickter & Sestir, 2006) demonstrou que o consumo de álcool prejudica de forma significativa a habilidade para

²⁶ Em outro estudo, Richeson (Richeson *et al.*, 2003) apresentou neuroimagens consistentes com essa interpretação.

regular vieses implícitos e, deste modo, inibir respostas racistas. Segundo os autores, esse resultado se deve ao enfraquecimento do controle cognitivo provocado pelo álcool²⁷.

Portanto, há toda uma linha de pesquisa elaborada a partir do paradigma da dupla cognição, que fornece muitos indícios de que pelo menos em relação a certos comportamentos considerados moralmente relevantes processos automáticos acabam cedendo a processos cognitivos mais lentos e reflexivos (Mallon & Nichols, 2010, p. 300-2).

Como mencionei anteriormente, Haidt apresenta inúmeros casos nos quais as pessoas condenam moralmente uma conduta, mas não conseguem encontrar justificativas para o julgamento — *moral dumbfounding* (Haidt, 2001). Os resultados apresentados por Haidt são frequentemente empregados por aqueles que pretendem demonstrar que os processos conscientes de reflexão não estão conectados com os julgamentos morais (*e.g.*, Dwyer, 2009).

No entanto, esses resultados não demonstram necessariamente que o julgamento moral não foi produzido por um processo de raciocínio consciente. Em casos assim é importante fazermos uma distinção entre justificativa e explicação. As pesquisas de Haidt demonstram que os participantes frequentemente falham quando são instigados a fornecer uma justificativa plausível para os seus julgamentos. Contudo, essas falhas não implicam uma incapacidade para explicar esses mesmos julgamentos. Quando um participante diz que dois irmãos fazerem sexo é errado porque é incesto, ele não oferece uma justificativa sobre por que devemos considerar o incesto errado, mas pode estar fornecendo uma explicação acurada sobre a origem de seu julgamento. As pessoas têm consciência sobre a existência de uma regra que proíbe o incesto. Portanto, quando elas dizem que ‘é errado porque é incesto’, elas podem estar identificando uma norma acessível à consciência que foi determinante na produção do julgamento (Mallon & Nichols, 2011, p. 285).

Ademais, embora as pessoas não sejam capazes de fornecer justificativas para seus julgamentos em algumas situações, elas são perfeitamente capazes em outras. Em um estudo conduzido por Cushman e seus colegas (Cushman et al., 2006), muitos participantes foram capazes de justificar os seus julgamentos com base na distinção entre ação e omissão. O fato mais relevante é que os julgamentos feitos previamente por aqueles que foram capazes de articular esse princípio realmente demonstravam uma maior utilização da distinção entre ação e omissão. Essa ligação entre ter um princípio elaborado conscientemente e utilizá-lo nos julgamentos indica que o raciocínio desempenha algum papel no processo de elaboração de julgamentos morais (Cushman, Young & Greene, 2010, p. 60).

²⁷ Para uma discussão mais aprofundada sobre esses estudos, ver Nosek *et al.* (2007) e Kelly, Machery & Mallon (2010).

Ainda, em certas circunstâncias, as pessoas alteram os seus julgamentos para amenizar as inconsistências de suas intuições iniciais (Lombrozo, 2009) ou quando são incentivadas a raciocinar sobre o assunto (Pizarro, Uhlmann, & Bloom, 2003, p. 657-8). Outros estudos de neuroimagem (Greene *et al.*, 2004) e de tempo de reação (Greene *et al.*, 2008)²⁸ associam processos de raciocínio aos julgamentos morais. Pelos menos em algumas oportunidades o compromisso consciente que temos com certas normas morais influencia as nossas deliberações sobre o que é certo ou errado (Cushman, Young & Greene, 2010, p. 47-8).

Todos esses resultados que indicam a importância do raciocínio para os julgamentos morais contradizem a posição não cognitivista clássica. Como procurei demonstrar por meio do argumento das expressões embutidas, é difícil explicar a partir de posições não cognitivistas esse tipo de raciocínio e as discussões morais travadas a partir dele.

As discussões sobre moralidade indicam que algo mais do que emoções está envolvido. Emoções podem desempenhar um papel importante na origem dos julgamentos morais, mas quando dizemos que alguma conduta é errada não queremos dizer apenas que não gostamos dela. Nossos julgamentos envolvem a formação de crenças e o conjunto dessas crenças acaba formando uma teoria normativa. Quando discutimos sobre a moralidade, estamos discutindo sobre o conteúdo e as implicações de uma teoria normativa que é em alguma medida compartilhada dentro de um grupo (Nichols, 2008, p. 268).

Como disse no começo desse capítulo, Darwin já percebia que os julgamentos morais não podem ser confundidos com emoções ou com instintos, pois requerem uma série de capacidades intelectuais. Não há somente um processo psicológico responsável pela produção de julgamentos morais. Eles são resultados de interações complexas entre diversos sistemas. Portanto, uma abordagem sobre eles não deve se restringir a debater se os julgamentos morais são fruto de processos cognitivos ou de ‘affects’; de raciocínio consciente ou de intuições (Cushman, Young & Greene, 2010, p. 66-8).

Dentre as diversas capacidades intelectuais envolvidas na moralidade, destaca-se a autoconsciência. Para julgarmos normativamente uma ação que nós mesmos praticamos precisamos ter a capacidade de identificar os motivos que nos levaram a praticá-la e de refletir sobre eles, ou seja, de compará-los com outros motivos que nos poderiam ter motivado a agir de forma diferente. Pensar de maneira normativa exige que tenhamos consciência dos fundamentos de nossas crenças e ações, e isso implica a autoconsciência, pois envolve a capacidade de nos

²⁸ Nesse estudo Greene e seus colegas verificaram que a manipulação da carga cognitiva dos testes interferia de modo seletivo com os julgamentos morais utilitários. Eles concluíram que esse resultado é consistente com a teoria que associa julgamentos morais utilitários a processos cognitivos controlados e julgamentos não utilitários a respostas emocionais automáticas.

identificarmos como os sujeitos de nossas representações mentais. Um ser sem autoconsciência pode estar consciente da existência de um objeto que ele deseja e agir de acordo com essa informação. No entanto um ser com autoconsciência, além de estar consciente da existência do objeto está também consciente do fato de que *ele* deseja o objeto. Ele não pensa apenas no objeto que deseja, mas também nos seus próprios desejos que o inclinam a agir de determinada forma. Esta autoconsciência a respeito dos motivos garante uma distância reflexiva que possibilita ao sujeito questionar os seus próprios motivos (Korsgaard, 2006, p. 112-116). Através da autoconsciência o indivíduo é capaz de se colocar na posição de espectador de seus próprios desejos. A partir dessa posição, ele pode comparar as suas ações no presente às passadas e aprová-las ou desaprová-las. A autoconsciência possibilita a existência de atitudes de segunda ordem, em que o sujeito se torna capaz de ter desejos sobre os seus próprios desejos e crenças sobre as suas próprias crenças; de pensar que teria sido melhor se tivesse agido de acordo com outro desejo (Baker, 2013, p. 191-192; Darwin, 2009, p. 73-74)²⁹.

Esse tipo de reflexão requer outras capacidades, como a memória e a linguagem. A capacidade de refletir sobre os motivos que nos levaram a agir de determinada forma depende da capacidade de nos lembrarmos de nossas ações, dos seus motivos e dos resultados produzidos: a lembrança de que uma ação provocou um sentimento de insatisfação pode, *e.g.*, fornecer subsídios para o juízo de que *deveríamos* ter agido de outra forma (Darwin, 2009, p. 70-72). O agente que dispõe de autoconsciência e de uma capacidade para se recordar de experiências passadas também dispõe da capacidade para pensar normativamente (Baker, 2013, p. 191-192), ainda que em um sentido puramente prudencial.

A linguagem, por sua vez, também é condição necessária para autoconsciência, tendo em vista que a existência desta depende da capacidade do sujeito ter um conceito de si mesmo — *self-concept* (Baker, 2013, p. 135). Além de possibilitar a reflexão consciente sobre os motivos de nossas ações e a formulação da crença de que deveríamos ter agido de outra forma, a linguagem permite também que compartilhem essas reflexões e crenças com os outros indivíduos de nosso grupo. Como indica o modelo social-intuicionista desenvolvido por Haidt (2012), esse tipo de interação social é outro importante elemento na conformação da moralidade.

Além da autoconsciência, da linguagem e da memória, a habilidade de atribuir estados mentais a outras pessoas e de compreender que esses estados mentais podem ser diferentes daqueles que possuímos também se vincula diretamente à produção de julgamentos morais. Animais em geral são capazes de considerar os comportamentos passados de outros indivíduos a fim de orientar as suas condutas futuras. No entanto, seres humanos não se satisfazem com

²⁹ Sobre a questão do papel da autoconsciência, ver também Damásio (2010).

simples descrições de comportamento. Nós procuramos compreendê-los postulando a existência de uma mente por trás deles; buscamos explicar *por que* o indivíduo se comportou de determinada forma, e essas explicações envolvem suposições sobre crenças e objetivos (Fitch, 2010, p. 135-40; Dennett, 1989, p. 17).

Assim como ocorre com relação à autoconsciência, a nossa habilidade de atribuir estados mentais também nos torna seres capazes de ter atitudes de segunda ordem. A capacidade que temos de ter crenças sobre crenças é fundamental para explicar muitos dos nossos julgamentos morais, pois, como regra, julgamos moralmente apenas os seres que acreditamos serem capazes de agir intencionalmente: não dizemos que uma telha agiu errado ao descolar-se do telhado e atingir a cabeça de alguém. A seguir, ao analisarmos o desenvolvimento e a evolução da moralidade, é recomendável ter em mente a forma como essas capacidades se relacionam.

4 Teoria da gramática moral universal

A versão mais conhecida e discutida de nativismo moral emprega a linguagem como um modelo para a cognição moral. A partir dos trabalhos de Noam Chomsky, passou a ter grande aceitação na área da linguística a ideia de que o cérebro possui uma estrutura dedicada ao aprendizado da linguagem (*language organ, innate language module, language faculty*). Na visão de Chomsky, essa faculdade, presente antes de qualquer aprendizado, envolve o conhecimento de várias regras, restrições e princípios linguísticos, tornando possível o aprendizado da linguagem.

Esse conhecimento constituiria um estado inicial, o qual, por meio da interação com as experiências linguísticas a que somos expostos durante a infância, resultaria na aquisição de um novo corpo de conhecimento linguístico, ou seja, na aquisição de uma língua específica — Chinês, Inglês, etc. A faculdade da linguagem poderia ser pensada como um cardápio de opções gramaticais básicas. Uma dessas opções seria, por exemplo, a ordem na qual sujeito, verbo e objeto são colocados em uma sentença. Na maioria das línguas humanas a ordem adotada é sujeito-verbo-objeto ou sujeito-objeto-verbo, quase nenhuma língua emprega a ordem objeto-sujeito-verbo. Esse cardápio também serviria como uma restrição, definindo quais linguagens somos capazes de aprender. A existência dessa restrição implica que o conjunto das línguas humanas naturalmente possíveis não coincide com o conjunto das línguas logicamente possíveis. Assim como não somos capazes de utilizar a ecolocalização, como fazem os morcegos, porque não temos o maquinário necessário, também não podemos, pelo mesmo motivo, aprender certas línguas possíveis (Hauser, 2007, p. 73-4; Dwyer, Huebner & Hauser, 2010, p. 489). Chomsky denominou esse cardápio ‘Gramática Universal.’ Ele estaria disponível a todas as crianças, e as experiências culturais delas determinariam as opções escolhidas, sendo que diferentes grupos selecionariam diferentes opções, ou parâmetros (Sripada, 2008, p. 325; Hauser, 2007, p. 38). Portanto, as regras não seriam aprendidas pela experiência. A experiência serviria apenas para determinar qual variação dentre as regras disponíveis será adotada (Prinz, 2012).

Chomsky demonstrou que a aquisição da linguagem é uma atividade extremamente complexa e não pode ser explicada simplesmente pelas nossas experiências. Para conseguirmos reconhecer uma sentença construída adequadamente, utilizamos princípios que estão além de nossa consciência, e a execução dessa tarefa exige construções teóricas que superam em muito o nosso conhecimento gramatical explícito (Hauser *et al.*, 2008, p. 116-7). Chomsky chama atenção para o fato de que quando uma criança consegue dominar a sua língua materna, antes

mesmo de receber qualquer instrução formal, ela já é capaz de fazer diversos julgamentos intuitivos³⁰ envolvendo as propriedades dessa língua, tais como se uma sequência aleatória de sons constitui uma sentença gramatical, se uma expressão é ambígua, se expressões arbitrárias formam uma rima, uma paráfrase ou são contraditórias. Essas e outras capacidades linguísticas seriam inexplicáveis caso a criança não possuísse um conhecimento tácito sobre a gramática de sua língua (Mikhail, 2011, p. 4).

A possibilidade de formularmos e de interpretarmos uma infinidade de sentenças a partir de um léxico limitado exige que elas sejam construídas com base em certos princípios e regras generativos. A habilidade de um falante compreender e produzir um conjunto infinito de novas sentenças depende da existência de um conhecimento do sistema gramatical da língua em questão. O conhecimento dessa gramática generativa extrapola em muito as normas presentes nas gramáticas pedagógicas. As gramáticas elaboradas com o intuito de auxiliar no aprendizado de uma língua qualquer dedicam-se a expor por meio de exemplos as particularidades e as exceções às regras dessa língua. Logo, elas não abordam os aspectos universais da linguagem (Chomsky, 1965).

Vários autores tentaram aplicar o argumento de Chomsky à aquisição da moralidade. Marc Hauser, John Mikhail e Susan Dwyer, entre outros, procuraram traçar um paralelo entre as cognições moral e linguística³¹, introduzindo a ideia de que haveria uma Gramática Moral Universal — GMU (Dwyer, 2007; Hauser, 2007; Mikhail 2011; Hauser *et al.*, 2008). A nossa capacidade para a moralidade envolveria princípios e construções conceituais que vão muito além do que pode ser explicado pelas normas e padrões a que somos expostos em nossa vida cotidiana. Os julgamentos morais que fazemos sobre uma infinidade de ações e omissões seriam resultado de um sistema complexo de regras, conceitos e princípios — inacessíveis em grande medida à nossa consciência — responsável por produzir representações mentais de vários tipos (Mikhail, 2011, p. 16). A moralidade de cada indivíduo seria o resultado de um conjunto universal de princípios e dos ajustes de parâmetros definidos em função das interações culturais (Hauser *et al.*, 2008, p. 122). De modo semelhante ao que ocorre com a linguagem, os princípios que compõem a GMU também limitariam quais sistemas morais somos capazes de adotar (Hauser, 2007, p. 73-4; Dwyer, Huebner & Hauser, 2010, p. 490).

De fato, há certas semelhanças entre línguas naturais – Inglês, Japonês, Português – e normas morais. Em ambos os domínios há um complexo padrão de características comuns e de

³⁰ Empregamos o conceito de intuição para nos referirmos aos julgamentos feitos rapidamente sem a aplicação consciente de qualquer princípio.

³¹ John Rawls teria sido o primeiro a traçar esse paralelo com as ideias de Chomsky (Hauser *et al.*, 2008). No entanto, a analogia entre moral e linguagem é muito mais antiga. Adam Smith, *e. g.*, já a utilizava (Smith, 1759/2014, p. 447).

diferenças entre grupos humanos (Sripada, 2008, p. 324-5). A capacidade para fazer julgamentos morais desenvolve-se cedo e de forma robusta, em contraste com outras capacidades como a de aprender a ler e a fazer cálculos, as quais exigem treinamento extensivo e podem perder-se facilmente (Sterelny, 2012). No entanto, como mostrarei nas próximas seções, há também grandes diferenças que colocam em dúvida a utilidade teórica desse tipo de analogia.

4.1 Argumentos da pobreza de estímulos e da projeção

Os principais argumentos utilizados para sustentar a existência de uma Gramática Universal focam-se na pobreza de estímulos (*poverty of the stimulus argument*). De acordo com esse argumento, a melhor explicação para o fato de crianças adquirirem os princípios que as capacitam a elaborar um número ilimitado de sentenças é a hipótese de que o cérebro humano dispõe de um sistema determinado geneticamente para o aprendizado da linguagem (Mikhail, 2011, p. 17; Dupoux & Jacob, 2007, p. 373). Esse sistema possibilitaria que crianças aprendessem as suas línguas maternas a partir de um conjunto finito e fragmentário de exemplos. Conforme esse tipo de argumento, há um problema em explicar como determinada capacidade cognitiva é adquirida quando há uma lacuna entre a complexidade do objeto do aprendizado e os recursos disponíveis ao aprendiz. A existência dessa lacuna serve então como evidência de que deve haver algum tipo de estrutura anterior a qualquer processo de aprendizado responsável pelo preenchimento desse vazio (Sripada, 2008, p. 325-6). O fato de um determinado traço desenvolver-se de maneira confiável mesmo perante ambientes variáveis e com apenas poucos estímulos seria evidência de que o traço em questão é ‘inato’, no sentido desenvolvimental do termo. O fato de um traço ser canalizado, por sua vez, é frequentemente interpretado como evidência de que ele é produto da evolução e, em alguns casos, de que ele é uma adaptação (Machery & Mallon, 2010).

Com relação ao aprendizado da linguagem, os estímulos são pobres no sentido de que eles não contêm qualquer evidência relevante de certos princípios adquiridos pelas crianças. Qual estímulo, *e.g.*, poderia explicar o fato de crianças com idades entre 3 e 5 anos serem capazes de identificar que na frase “I wonder who the men expected to see them” o pronome “them” pode referir-se a “the men”, enquanto na frase “the men expected to see them” isso não é possível? ou que nas sentenças “Sally squirted paint on the wall” e “Sally squirted the wall with paint” é possível inverter os objetos direto e indireto, enquanto a inversão na sentença “Peggy poured paint on the floor” resulta na sentença não gramatical “Peggy poured the floor with paint”? (Prinz, 2012)? Embora sejamos capazes, ainda na infância, de dominar a aplicação das

regras que explicam essas distinções, Chomsky salienta que elas não estão presentes nas gramáticas pedagógicas e sequer eram conhecidas pelos linguistas há algumas décadas (1986, p. 8).

Além disso, durante o período de aprendizado, as crianças têm acesso apenas a uma base de informações degenerada, no sentido de que elas são expostas a muitas sentenças gramaticalmente incorretas. As sentenças que elas ouvem apresentam muitas expressões que não são bem formadas, em razão de mudanças de planos no meio da fala, de finais de frase que não se coadunam com o começo, de frases que não são concluídas, de atos falhos, *etc.* Ademais, um estudo feito por Roger Brown e Camille Hanlon indica que pais frequentemente aprovam sentenças incorretas produzidas por seus filhos (Brown & Hanlon, 1970). Apesar disso, as crianças acabam aprendendo a produzir sentenças corretas. Se elas estivessem apenas fazendo extrapolações a partir do que ouvem, esse aprendizado não seria explicável (Mikhail, 2011, p. 70).

Defensores da GMU defendem que o comportamento e o discurso moral dos adultos aos quais as crianças são expostas não incluiriam informações suficientes para permitir a aquisição de competência moral se não houvesse mecanismos de aprendizado dedicados exclusivamente ao domínio moral (Prinz, 2008, p. 430; Joyce, 2006, p. 137). Saber se os estímulos são suficientes para explicar a aquisição de princípios morais depende, entre outras coisas, de sabermos se esses princípios são invocados de forma expressa ou estão meramente implícitos nos nossos julgamentos. Muitos princípios linguísticos são enigmáticos e não são conhecidos pelos falantes da língua. Essa situação também poderia ocorrer em relação a certos princípios morais empregados pelas pessoas em geral sem que elas tenham consciência disso ou sejam capazes de ensiná-los às crianças (Harman, 2008, p. 347). Um possível exemplo disso seria o caso do princípio do duplo efeito, que será analisado mais adiante.

No caso da linguagem, a existência de uma lacuna entre o objeto a ser aprendido e os recursos disponíveis para o aprendizado é bastante plausível. O aprendizado da linguagem é necessariamente uma tarefa indutiva: quem aprende deve identificar regras gerais com base em exemplos particulares da regra. Pesquisas recentes utilizando modelos de aprendizagem de linguagem sugerem que aprendizados indutivos dessa espécie podem até mesmo ser impossíveis se não houver fortes restrições no campo de hipóteses (Sripada, 2008, p. 327-8).

No entanto, ao contrário do que ocorre com a linguagem, o aprendizado de normas morais não precisa ser indutivo. As crianças são instruídas expressamente sobre as normas morais, religiosas e jurídicas.

No âmbito da linguagem as crianças não recebem informações negativas explícitas. Certas sequências não gramaticais não são produzidas e rotuladas como desviantes, elas simplesmente não são produzidas. Assim, é necessário explicar o que restringe as possibilidades gramaticais. No que diz respeito à moralidade, atos proibidos são produzidos e descritos, a proibição é explicitada e há instrução intensiva (Sterelny, 2010, p. 290-1). As crianças transgridem as normas diariamente, elas destroem coisas e muitas vezes são violentas e egoístas (Prinz, 2012). Essas atitudes são repreendidas com frequência; pais e professores constantemente dizem às crianças o que não deve ser feito (Nichols, 2005, p. 358; Minton, Kagan & Levine, 1971). O psicólogo Martin L. Hoffmann, com base em diversos estudos empíricos, afirma que crianças entre 2 e 10 anos de idade recebem cerca de 50 advertências diárias de seus pais, *i.e.*, mais de 15.000 em um ano (Hoffman, 2001, p. 141).

Hauser, Dwyer e Huebner reconhecem que as crianças são repreendidas, mas segundo eles essas instruções não são suficientes para explicar as regras morais sofisticadas que são encontradas nos julgamentos morais feitos pelas crianças. Em especial, os imperativos que elas ouvem ao serem ensinadas – ‘sempre cumpram as suas promessas’ – não seriam suficientes para explicar como as crianças são capazes de identificar os casos que constituem exceções a essas regras (Dwyer, Huebner & Hauser; 2010, p. 492).

No entanto, as instruções explícitas não são tão grosseiras como Dwyer, Huebner e Hauser pretendem. Os pais normalmente oferecem explicações para as suas instruções. Eles recorrem a diferentes razões e normas para justificar por que a criança deve alterar o seu comportamento. Além disso, em muitas das ocasiões os pais procuram induzir a criança a adotar a perspectiva da pessoa ofendida, indicando o sofrimento provocado e deixando claro que a criança foi a responsável por esse dano (Hoffman, 2001, p. 143). As instruções explícitas podem ser empregadas de diferentes formas, dependendo do contexto e do tipo de transgressão (Nilsen, 2013, p. 193-4).

Além de instruções explícitas, as crianças recebem outros estímulos. Os defensores da GMU adotam um entendimento muito restrito acerca do que pode ser considerado como um ‘input’ relevante do ponto de vista moral (Nilsen, 2013, p. 192-3). As histórias, mitos e canções às quais as crianças têm acesso são repletos de lições sobre a moralidade, como observa Sterelny:

The narrative life of a community—the stock of stories, songs, myths, and tales to which children are exposed—is full of information about what actions are to be admired, and which are to be deplored. Young children’s stories include many moral fables: stories of virtue, of right action and motivation rewarded, of vice punished. Their narrative world is richly populated with moral examples (Sterelny, 2010, p. 288-9).

Além do mais, a vida das crianças não é passiva. Elas não ficam apenas escutando histórias e assistindo ao que os outros fazem. Elas agem e enfrentam conflitos com outras crianças envolvendo distribuição de recursos e produção de danos. Possivelmente as crianças não seriam capazes de aprender as normas sobre distribuição de bens apenas pela observação do comportamento de seus parentes próximos, mas elas participam com frequência de discussões nas quais se busca encontrar uma regra apropriada para solucionar esse tipo de conflito — *e.g.*, eu divido e você escolhe. Interações como essas, que compõem o processo de aprendizado, são carregadas de aspectos ligados à moralidade (Sterelny, 2010, p. 288-91).

A ênfase nesses processos explícitos e implícitos de aprendizado serve para demonstrar a importância que a construção de nichos tem no desenvolvimento da moralidade nas crianças. Os pais e os outros membros da sociedade organizam o ambiente das crianças de tal maneira que ele ofereça as informações que elas precisam. O ambiente no qual elas se desenvolvem é moldado a fim de garantir que o aprendizado moral, por meio de tentativas e erros, ocorra de maneira segura. Esse ambiente construído culturalmente é composto por diversos brinquedos, ferramentas, jogos, histórias e exemplos. Os pais, segundo Sterelny, funcionam como verdadeiros engenheiros na construção do ambiente de aprendizado, a fim de garantir que a transmissão social das informações normativas ocorra com sucesso de uma geração para outra. Essa organização do ambiente no qual as crianças se desenvolvem explicaria o desenvolvimento robusto da moralidade (Sterelny, 2010, p. 290-1). O ambiente no qual as crianças se desenvolvem não seria, portanto, pobre em estímulos. Pelo contrário, ele seria saturado com múltiplas experiências particulares envolvendo valorações normativas, as quais funcionariam como *input* em um sistema de aprendizado capaz de reconhecer padrões (Sterelny, 2012). A ênfase na construção do nicho ontogênico está atrelada à ideia de que os organismos constroem os seus próprios ambientes e que essa construção funciona como um mecanismo de herança. O organismo que está se desenvolvendo herda o nicho ontogênico assim como ele herda o seu genoma, ainda que segundo meios distintos de transmissão (Stotz, 2010, p. 485 e 488).

Em suma, não é difícil demonstrar que o ambiente no qual as crianças se desenvolvem apresenta muitos estímulos morais, explícitos e implícitos. No entanto, isso não é suficiente para refutar o argumento da pobreza de estímulos. A refutação exige a demonstração de que as informações recebidas são suficientes para explicar a capacidade desenvolvida pela criança. Na área da linguagem, as crianças também recebem uma série de estímulos. A existência da pobreza de estímulos não se relaciona apenas com a quantidade de estímulos, mas também com a complexidade daquilo que é aprendido. Envolve demonstrar que os dados disponíveis não são suficientes para explicar a aprendizagem.

No âmbito da linguagem, a complexidade da capacidade adquirida é ilustrada pelo paradoxo da aquisição da linguagem, formulado por Jackendoff (1994). Collins o resume da seguinte maneira:

The professional communities of linguists and psychologists have all the data they desire on, say, English, certainly much more than any child has ever enjoyed. They also pool their mature intellects, have equal data on many other languages, and design subtle experiments and conduct longitudinal studies over many years. Yet they still can't work out the principles, rules and concepts that constitute competence with English which the normal child masters by the time she is five! So, it is not so much that the child's data is poor according to some absolute standard (as if such a notion makes sense); rather, if the child were relying on data, then her acquisition of language would be miraculous. The acquisition ceases to be at all miraculous if we understand the child to possess innate resources which already determine most of what is to be acquired. On this view, the child's task is so 'easy' precisely because she isn't dependent on rich data, while the data inundated scientist's task is so hard precisely because he is trying to work out what the child already knows independent of data or general principles defined over it (2004, p. 505).

Resta aos defensores da GMU a tarefa de demonstrar que a capacidade para fazer julgamentos morais é tão complexa quanto a capacidade linguística. Como veremos adiante, o objeto do aprendizado no caso da moral parece ser muito mais simples do que no caso da linguagem. Normas morais não são tão obscuras quanto as regras recursivas e as ramificadas regras estruturais das gramáticas humanas. O desenvolvimento da capacidade moral requer o aprendizado de um conjunto de normas mais concretas, *e.g.*, compartilhe os seus brinquedos, não bata em outras crianças, respeite os mais velhos, *etc.* (Sripada, 2008, p. 328).

No entanto, há ainda outro argumento que reforça o argumento da pobreza de estímulo: como podemos explicar, a partir de um número finito de experiências, o desenvolvimento de uma capacidade de julgar infinitos casos?

Assim como os linguistas, os teóricos que procuram explicar a moralidade também enfrentariam o problema da projeção (*projection problem*), *i.e.*, o problema de explicar como indivíduos são capazes de aplicar o seu conhecimento moral a casos novos diferentes de todos os precedentes (Mikhail, 2011, p. 30). Um falante já maduro de alguma língua encontrou apenas um número limitado de sentenças ao longo de sua vida. No entanto, a partir dessa experiência limitada, ele é capaz de construir um número infinito de sentenças que são compreensíveis para os outros falantes da língua. Ademais, ele é capaz também de julgar intuitivamente vários aspectos gramaticais dessas possíveis sentenças e várias relações entre as sentenças e seus componentes. Ele sabe também identificar se uma sequência qualquer de sons constitui uma sentença gramaticalmente correta, se ela é ambígua, se é uma paráfrase de outra sentença, *etc.* (Chomsky 1975, p. 61–62). Alguém capaz de falar inglês, por exemplo, reconhece que: a) a sentença “colorless green ideas sleep furiously”, embora não faça sentido, é gramatical,

enquanto a sentença “furiously sleep ideas green colorless” não o é; b) a sentença “the little boys’ camp is on fire” é ambígua; c) na sentença “I wonder who the men expected to see them” o pronome “them” pode referir-se a “the men”, mas na sentença “the men expected to see them” não o pode; d) a questão “is the man who is here tall?” está formulada corretamente, enquanto a sentença “is the man who here is tall?” não está. Essa pessoa será capaz de fazer essas distinções ainda que não as saiba explicar coerentemente e nunca tenha escutado tais sentenças (Mikhail, 2011, p. 45).

Desta forma, um falante que domine uma língua é capaz de produzir um número potencialmente infinito de julgamentos a respeito das propriedades e relações entre expressões nessa língua. Tendo em vista que a capacidade de armazenamento de nosso cérebro é limitada, não é possível que todas essas expressões estejam armazenadas nele individualmente. Portanto, deve haver uma gramática — *i.e.*, um programa ou uma fórmula de algum tipo — por meio da qual possamos, a partir de um léxico limitado e de padrões de frases, construir e interpretar um número ilimitado de expressões (Chomsky, 2002; Mikhail, 2011, p. 45-6; Dwyer, Huebner & Hauser, 2010, p. 488-9).

Os defensores da GMU propõem que algo semelhante ocorre nesse domínio: um indivíduo com o senso de justiça desenvolvido seria apto a fazer um número ilimitado de julgamentos morais intuitivos a respeito das propriedades morais de ações e de agentes. Nós enfrentamos cotidianamente casos que nunca havíamos examinado e ainda assim não temos dificuldades em avaliá-los (Dwyer, Huebner & Hauser; 2010, p. 489). Além disso, como regra, não há duas situações que sejam exatamente iguais. Cada uma delas combina de forma única agentes, atos, eventos e circunstâncias (Mikhail, 2011, p. 47-8). De novo, considerando a capacidade limitada de armazenamento do cérebro e a capacidade ilimitada de fazer julgamentos, segue que a faculdade moral é mais do que uma simples lista. Haveria uma gramática moral, ou seja, um sistema cognitivo composto por princípios e regras capazes de produzir um número ilimitado de representações mentais que resultariam nos julgamentos morais (Mikhail, 2011, p. 46 e 72).

A rejeição de nativistas como Mikhail e Hauser ao particularismo moral³² — ou seja, à ideia de que não existem princípios morais — remonta a uma visão presente na filosofia da ciência, segundo a qual há simplicidade por trás dos fenômenos, os quais poderiam ser explicados por um pequeno conjunto de princípios. David Hume já expressava claramente essa ideia em relação à moral:

It may now be asked in general, concerning this pain or pleasure, that distinguishes moral good and evil, from what principles is it derived, and whence does it arise in the

³² Sobre o Particularismo Moral, ver o artigo ‘Moral Particularism’ (Dancy, 2013).

human mind? To this I reply, first, that it is absurd to imagine, that in every particular instance, these sentiments are produced by an original quality and primary constitution. For as the number of our duties is, in a manner, infinite, it is impossible that our original instincts should extend to each of them, and from our very first infancy impress on the human mind all that multitude of precepts, which are contained in the compleatest system of ethics. Such a method of proceeding is not conformable to the usual maxims, by which nature is conducted, where a few principles produce all that variety we observe in the universe, and everything is carried on in the easiest and most simple manner. It is necessary, therefore, to abridge these primary impulses, and find some more general principles, upon which all our notions of morals are founded (Hume, 2009/1740, p. 720-1).

Assim como na linguagem, as pessoas podem tentar explicitar os princípios que empregaram para chegar ao julgamento a respeito da moralidade de certa conduta, mas esses princípios podem não ser os princípios que realmente explicam a competência moral (Daniels, 1980, p. 22-3), ou seja, os princípios acessíveis à consciência não seriam suficientes para explicar coerentemente os julgamentos morais. Além disso, experiências como aquelas envolvendo ‘trolley problems’ indicam que somos capazes de julgar com rapidez mesmo em situações com as quais não temos qualquer familiaridade. Esse caráter intuitivo dos julgamentos sugeriria que eles ocorrem segundo processos computacionais inacessíveis à nossa consciência.

No entanto, embora os proponentes da GMU aleguem que uma das vantagens dessa teoria é a simplicidade, os seus críticos afirmam justamente o contrário, ou seja, que ela é muito mais complicada do que o necessário para explicar o fenômeno a que se propõe.

Nas ciências em geral, há o objetivo de explicar o maior número possível de fenômenos com um número pequeno de axiomas. A teoria da GMU se alia a esse objetivo na medida em que busca explicar a moralidade recorrendo a um conjunto limitado de princípios. Contudo, ao recorrer à analogia linguística, ela adota um modelo bastante complicado de princípios. Talvez os julgamentos de aprovação e reprovação morais sejam mais semelhantes aos que fazemos quando experimentamos alguma comida do que quando julgamos a gramaticalidade de uma sentença. Para conseguirmos explicar como somos capazes de interpretar frases completamente novas a partir dos sons que ouvimos, talvez seja de fato necessário apelarmos para um conjunto complexo de princípios, pois há uma grande dificuldade em explicar a reversibilidade presente na linguagem, *i.e.*, como é possível que alguém possa produzir um conjunto novo de sons a partir de representações semânticas (significados) e outra pessoa possa reconstruir essas representações semânticas com base nesse conjunto de sons nunca antes ouvido. Quando um falante nativo faz um julgamento a respeito da gramaticalidade de uma sentença, ele não julga apenas se aquela sentença é correta ou incorreta, ele também busca compreender (reconstruir) o que a outra pessoa quis dizer. Como afirmam Dupoux e Jacob, no âmbito dos julgamentos morais não há nada semelhante à reversibilidade presente na linguagem, enquanto esta constitui um sistema generativo, a moralidade é apenas um sistema avaliativo: “it makes sense to map the

structural description of an agent's act onto a valence. But it would make little (if any) sense to generate a structural description of an act in response to the representation of the valence of that act" (Dupoux & Jacob, 2007, p. 376).

Além disso, a teoria da GMU também não é simples no sentido de que ela postula a existência de faculdades específicas e de módulos especializados, cuja existência, como veremos no próximo capítulo, é bastante controversa. Nesse contexto, teorias construtivistas baseadas no reconhecimento de padrões têm a vantagem da simplicidade sobre a teoria GMU, na medida em que procuram explicar o aprendizado da moralidade através de mecanismos que são empregados em outras modalidades de conhecimento. O primeiro passo para defender a existência de uma capacidade específica para o aprendizado moral seria demonstrar a impossibilidade de ele acontecer por meio de sistemas mais gerais, como o responsável pelo reconhecimento de padrões. Cabe aos teóricos da GMU o ônus de demonstrar a impossibilidade de esclarecer os nossos juízos morais através de teorias mais simples. Ninguém duvida que temos uma capacidade para reconhecer padrões. Logo, antes de propormos a existência de um sistema especializado, deveríamos explicar por que razão um sistema geral não é capaz de fornecer uma explicação adequada (Prinz, 2012).

4.2 'Trolley problems' e princípio do duplo efeito

Para testar se uma regra gramatical qualquer está presente na gramática generativa utilizada pelos falantes de alguma língua, linguistas elaboram problemas envolvendo julgamentos de aceitabilidade e os confrontam com as intuições dos participantes; *i.e.*, eles elaboram diferentes sentenças e perguntam a falantes comuns daquela língua se elas são aceitáveis ou apresentam algum problema. Por exemplo, se um princípio gramatical proposto implica que a sentença "John is having run" está bem formulada, mas essa sentença conflita com as intuições dos participantes, então os resultados pesam contra a aceitação desse princípio (Roedder & Harman, 2016, p. 4-5).

Nesses testes, as respostas comuns costumam ser convergentes, e os participantes mostram-se confiantes quanto aos seus julgamentos, embora não saibam explicar os princípios computacionais que produziram o julgamento de gramaticalidade. Teóricos da GMU costumam seguir esse modelo e elaborar experiências compostas por dilemas morais para tentar identificar as regras e princípios que são capazes de capturar as intuições dos entrevistados, identificando dessa forma as características da faculdade moral. O objetivo dessas experiências é compreender com que fundamento fazemos a distinção entre ações permitidas e proibidas. A hipótese é que a distinção é fornecida por uma gramática moral (Mikhail, 2011, p. 89). Adeptos da GMU

defendem que as respostas apresentadas aos problemas propostos costumam apresentar as mesmas características dos julgamentos a respeito da gramaticalidade de uma sentença, *i.e.*, os indivíduos respondem aos dilemas de modo rápido, confiante e convergente. A ideia é que esses julgamentos são produzidos por processos computacionais que podem ser identificados por meio de sucessivos testes envolvendo a manipulação sistemática dos traços relevantes dos dilemas morais (Dwyer, Huebner & Hauser; 2010, p. 493-4; Nichols, 2005, p. 359).

Marc Hauser e seus colegas utilizaram uma página da Internet para entrevistar milhares de pessoas sobre dilemas envolvendo bondes (*trolley problems*), originalmente formulados por Philippa Foot (Hauser *et al.*, 2008). Os resultados obtidos foram semelhantes entre os mais variados grupos demográficos; fatores como sexo, raça, idade, religião, nível de educação e vários outros fatores culturais mostraram-se irrelevantes. Pesquisas desse tipo foram reproduzidas com mais de 200 mil pessoas distribuídas em mais de 120 países (Mikhail, 2011, p. 38).

Com relação à versão mais simples do dilema, a grande maioria dos entrevistados respondeu que era moralmente permissível desviar o bonde para um trilho adjacente a fim de salvar a vida de cinco trabalhadores, ainda que essa conduta causasse a morte de um inocente. Contudo, eles não consideraram permissível lançar uma pessoa na frente do bonde a fim de salvar esses mesmos cinco trabalhadores. Apenas alguns dos entrevistados sabiam explicar por que uma das condutas era permissível e a outra não.

Segundo os pesquisadores, as pessoas tinham acesso consciente a alguns dos princípios relacionados aos seus julgamentos, mas não a outros. Na busca do melhor resultado possível — salvar o maior número de pessoas de um desastre iminente — os entrevistados consideravam aceitável tolerar consequências ruins que são previsíveis mas não são obtidas intencionalmente. No entanto, consideravam inaceitáveis condutas que intencionalmente provocavam uma consequência ruim para evitar um mal ainda maior. Isto é, os entrevistados julgavam de acordo com o princípio do duplo efeito, que pode ser formulado da seguinte forma: *it is worse to cause harm to someone (who has not consented to this) as (part of) your means to bringing about a greater good to others than to cause such harm as a side effect of doing something that will bring about a greater good* (Harman, 2008, p. 346).

Os resultados dessas pesquisas levaram Hauser e Mikhail a acreditar que os julgamentos morais são mediados por princípios inacessíveis à nossa consciência, pois os entrevistados eram incapazes de descrever de forma convincente os seus próprios processos de pensamento, e as respostas oferecidas eram incompatíveis com os julgamentos ou incoerentes (Hauser *et al.*, 2008, p. 133-4; Mikhail, 2011, p. 38 e p. 309-10). Havia uma dissociação entre julgamento e

justificação. Embora os julgamentos fossem os mesmos, as justificativas divergiam bastante (Mikhail, 2011, p. 84). Além disso, os entrevistados se mostravam intrigados em razão de suas intuições sobre casos tão semelhantes divergirem.

Além do princípio do duplo-efeito, nas diversas variações do dilema do bonde, outras normas teriam sido aplicadas pelos participantes, tais como a proibição de agressões e de homicídios intencionais e o ‘rescue principle’, que proíbe alguém de deixar de prevenir uma morte ou outro grave dano facilmente evitáveis, naquelas situações em que isso pode ser feito sem pôr em risco a própria segurança e sem violar preceitos morais fundamentais (Mikhail, 2011, p. 117).

Dwyer, Huebner e Hauser apontam para quatro importantes distinções que teriam sido identificadas a partir desses estudos empíricos: 1) danos causados por ação seriam piores do que aqueles causados por omissão; 2) danos causados como um meio para um bem maior são piores do que aqueles causados como efeito colateral de outra ação; 3) danos causados por contato físico direto são piores do que os causados por um intermediário não humano; 4) danos causados a alguém como meio de obter um bem maior são mais permissíveis se a pessoa que sofre o dano não fica em uma situação pior do que ficaria se o ato não fosse praticado – ‘Pareto improvement’ (Dwyer, Huebner & Hauser, 2010, p. 496-7).

Quando as pessoas pensam em características universais da moralidade elas não imaginam algo semelhante ao princípio do duplo-efeito. No entanto, isso não é um problema. As características universais da linguagem também não são o tipo de coisa que encontramos nas gramáticas escolares (Roedder & Harman, 2016, p. 12-13).

No campo da moral, não está claro como tais princípios como o do duplo efeito poderiam ser aprendidos, uma vez que eles raramente são articulados e não são generalizações óbvias de casos particulares. Se as pessoas em geral não os conhecem de forma consciente, elas não podem ensiná-los explicitamente. Se as crianças, apesar disso, os aprendem, elas parecem estar de alguma forma predispostas a isso (Roedder & Harman, 2010, p. 286-7). A possibilidade aventada é a de que eles estariam embutidos em nossa mente, inscritos na moralidade de cada indivíduo como parte da GMU (Harman, 2000, p. 225; 2008, p. 349; Sterelny, 2012).

Algumas outras explicações, além da aplicação do princípio do duplo-efeito, foram elaboradas para as respostas dadas aos dilemas envolvendo bondes. Alguns tentaram explicá-las por meio do imperativo categórico de Kant. Nesse sentido, empurrar alguém sobre os trilhos seria proibido, pois equivaleria a utilizar a pessoa empurrada como um meio para prevenir a morte de outras cinco, enquanto no caso em que o trem é apenas desviado ninguém estaria sendo

utilizado como um meio. Alguns dos dilemas formulados indicam que essa distinção é relevante para muitas pessoas.

Outros, como Greene (Greene & Haidt, 2002), consideram que a distinção entre danos produzidos de forma pessoal e de forma impessoal pode explicar as diferenças (*personal and impersonal harm*), pois os casos que produzem danos de forma pessoal seriam mais carregados de emoção e, por essa razão, considerados proibidos. Essas distinções podem ser testadas por meio de pequenas alterações nos dilemas. Por exemplo, faria alguma diferença se a pessoa tivesse simplesmente que apertar um botão, em vez de empurrar a outra com as próprias mãos sobre os trilhos (Hauser, 2007, p. 116-7)? Hauser acredita que não, mas não é isso que mostram os resultados das pesquisas. O número de pessoas que consideram proibido matar uma pessoa como meio para salvar outras cinco varia dependendo da forma como o dilema é formulado e da pessoalidade ou impessoalidade envolvida na ação. No seguinte cenário (Hauser, 2007, p. 32), inspirado pelos *trolley problems*, um paciente saudável está em um hospital esperando para fazer exames de rotina. Nesse mesmo hospital há outros 5 pacientes que necessitam receber transplantes de órgãos. Considerando que os órgãos do primeiro paciente podem salvar os outros cinco e que eles morrerão se não receberem os órgãos, seria moralmente permissível matar o primeiro paciente para extrair os seus órgãos? A esmagadora maioria das pessoas considera que não, e uma das explicações para isso pode ser o caráter pessoal dessa ação. A distinção de Greene não explica todos os resultados, mas ela indica um importante aspecto das decisões morais.

Por sua vez, Prinz defende que em casos como os dilemas envolvendo bondes, na decisão adotada pelos entrevistados não estão presentes regras ou princípios inconscientes. As pessoas aprendem que é errado matar por meio de casos paradigmáticos e elas são mais tolerantes em relação a condutas que se afastam desse paradigma, são menos salientes, ou são comparáveis a outras condutas que envolvem o mesmo tipo de morte mas perdas maiores (Prinz, 2013, p. 106).

A grande convergência de respostas encontrada no caso clássico do dilema envolvendo bondes pode ser explicada em razão de a solução considerada proibida moralmente — empurrar a pessoa obesa, derrubando-a de cima de uma passarela sobre os trilhos — combinar um conjunto de características normalmente consideradas reprováveis em diversas sociedades: a pessoa obesa está sendo empregada como um meio; o caso assemelha-se muito a casos paradigmáticos de homicídio e há um forte elemento de pessoalidade. A conduta de empurrar alguém, *e.g.*, é muito mais pessoal do que o ato de puxar uma alavanca, o que incentiva mais pessoas a considerá-la impermissível. Quando esses fatores são isolados, as respostas dos

participantes deixam de convergir para um resultado comum. Esse fato fica evidente quando analisamos os resultados de pesquisas envolvendo dois cenários denominados de “Loop Track” e de “Man-in-Front” por Mikhail (2011, p. 107-8). Mikhail descreve esses cenários da seguinte forma:

Loop track: Ned is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, giving the men time to escape. The heavy object is a man, standing on the side track with his back turned. Ned can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die. Is it morally permissible for Ned to throw the switch? (...)

Man-in-front: Oscar is standing next to a switch, which he can throw, that will temporarily turn the train onto a side track. There is a heavy object on the side track. If the train hits the object, the object will slow the train down, giving the men time to escape. There is a man standing on the side track in front of the heavy object with his back turned. Oscar can throw the switch, preventing the train from killing the men, but killing the man. Or he can refrain from doing this, letting the five die. Is it morally permissible for Oscar to throw the switch? (Mikhail, 2007, p.149).

A diferença essencial entre os dois cenários é que no primeiro deles a morte da pessoa localizada sobre o trilho adjacente é um meio necessário para evitar a morte das outras cinco. Enquanto no segundo caso, a morte daquela pessoa é apenas um efeito colateral do ato de desviar o trem. No primeiro cenário, 48% dos entrevistados consideraram permissível desviar o trem; no segundo, 62%. Embora os participantes tenham respondido de forma bastante dividida, Mikhail acredita que o princípio do duplo efeito foi observado, pois um número maior de participantes julgou impermissível a conduta que provocava um dano como meio para obter um bem maior.

A questão que se coloca, no entanto, é, em que medida podemos, a partir de resultados tão divididos, identificar supostos princípios que estariam por trás dos julgamentos morais. O fato de ser possível elaborar princípios, como o do duplo efeito, que sintetizam as intuições morais de uma grande parte da população não significa que essas intuições tenham sido produzidas com base nesses princípios. Nesse sentido, é importante distinguir entre uma abordagem externa e uma abordagem interna. A abordagem externa busca elaborar um conjunto de princípios que seja capaz de explicar a maioria das intuições manifestadas pelas pessoas, e é possível elaborar mais de um conjunto de princípios que cumpram igualmente bem essa tarefa. A abordagem interna, por outro lado, busca identificar aqueles princípios que são causalmente responsáveis pela produção das intuições observadas. Como Nichols salienta, a preocupação em relação aos *trolley problems* parece ser elaborar um conjunto de princípios que se coadune com a maioria das intuições, sem a preocupação de demonstrar que esses princípios estão envolvidos de fato na produção das intuições (Nichols, 2005, p. 360). O fato de que as intuições de muitas

peças não obedecem aos princípios acaba simplesmente sendo contornado pelo apelo à distinção entre competência e performance. Como veremos adiante, essa distinção apresenta inúmeros problemas quando aplicada à moralidade. Há uma grande dificuldade, *e.g.*, de identificar se determinadas emoções integram a competência moral ou são apenas fatores que prejudicam a performance.

Em vez de tentar explicar os resultados encontrados nessas pesquisas por meio de princípios específicos à moralidade, presentes em nossa mente antes de qualquer experiência, talvez seja possível explicá-los por meios de mecanismos gerais de aprendizado envolvidos na capacidade de reconhecer padrões. A mente humana é boa em reconhecer padrões, em perceber similaridades entre diferentes situações, e o exercício dessa capacidade resulta em julgamentos intuitivos sobre novos casos. Esse tipo de reconhecimento é feito a partir de mecanismos de aprendizado estatístico. Por meio desse tipo de mecanismo, animais são capazes de identificar regularidades como quais os melhores locais para encontrar alimentos ou quais objetos são comestíveis. Empregamos esse tipo de aprendizado, *e.g.*, quando aprendemos as sequências de movimentos musculares necessárias para executar determinada tarefa (Prinz, 2012).

O reconhecimento de padrões estende-se a infinitos novos casos e ocorre de maneira rápida e automática. Além disso, aquele que reconhece o padrão muitas vezes não consegue explicar qual foi a base utilizada para o julgamento. Um especialista em pássaros, por exemplo, pode reconhecer, em um relance de olhos, determinada espécie, mesmo que não saiba explicar exatamente o porquê (Sterelny, 2012; Sterelny, 2010, p. 287-8). Portanto, se as intuições morais fossem o resultado de um processo de reconhecimento de padrões não surpreenderia o fato de elas serem rápidas e de termos dificuldades em distinguir claramente as suas causas (Sterelny, 2010, p. 288-9).

Chomsky utiliza uma série de exemplos para defender que a gramática não pode ser aprendida pela captura de padrões lexicais, e que não aprendemos uma língua fazendo generalizações de casos familiares. Até agora não foi desenvolvido argumento semelhante em relação à cognição moral. Julgamentos morais intuitivos podem ser generalizações de casos exemplares: atos generosos são aqueles semelhantes a casos paradigmáticos de generosidade. Se os julgamentos morais puderem ser explicados pelo reconhecimento de padrões, os principais argumentos dos teóricos da GMU desabam, incluindo o argumento da pobreza de estímulos e da projeção. Para demonstrar que essa explicação não é suficiente, eles devem comprovar que as sutilezas e abstrações envolvidas nos julgamentos morais impedem que eles sejam simples generalizações de fatos observados. Até agora, os esforços empregados por Hauser e Mikhail para demonstrar as sutilezas envolvidas nos julgamentos morais voltaram-se a questões que não

são exclusivamente morais, mas que estão envolvidas na nossa capacidade geral de nos orientarmos no âmbito de convívio social, ou seja, a nossa capacidade para atribuir estados mentais e intenções, e para representar mentalmente as características de uma ação (Sterelny, 2012; Sterelny, 2010, p. 287-8). O caráter geral dessas capacidades contraria a pretensão de especificidade da faculdade moral.

Ainda que eventualmente o princípio do duplo efeito possa ser aplicado inconscientemente nas decisões, trata-se de uma exceção (Sterelny, 2012). Como regra, nós articulamos os princípios que motivam nossos julgamentos. Se o modelo da GMU fosse aplicável aos nossos julgamentos morais, teríamos respostas morais para situações específicas, mas não teríamos a capacidade ou a inclinação para oferecer qualquer explicação ou defesa dessas respostas. Casos como o da aplicação do princípio do duplo efeito seriam a regra, não a exceção, se houvesse uma gramática moral.

Na prática nós temos consciência dos princípios morais e articulamos nossos pensamentos morais, ainda que muitas vezes não haja correspondência entre esses pensamentos e as causas efetivas do nosso julgamento moral. Aqueles que adotam o modelo da GMU têm dificuldade em explicar a capacidade que possuímos para refletir e articular nossos julgamentos morais. Ao contrário do que ocorre com a moral, nós não temos acesso introspectivo à Gramática Universal, aos princípios sintáticos que usamos para organizar palavras e frases, e para decodificar a fala de outras pessoas (Sterelny, 2012).

Sterelny vê a relação entre princípios tácitos e princípios explícitos presentes na moralidade como algo muito semelhante ao que ocorre com outras habilidades adquiridas por meio de uma capacidade geral de aprendizado. Um artesão, por exemplo, possui bastante conhecimento explícito — ele pode explicar facilmente muitas das técnicas que ele utiliza — que coexiste com conhecimentos implícitos resultantes do hábito de exercer a profissão — *know-how* — e com uma capacidade de reconhecer padrões intuitivamente (Sterelny, 2012). Contudo, essa distinção entre conhecimento explícito e conhecimento implícito, tanto no âmbito moral quanto em relação à atividade do artesão, não é muito precisa: o artesão pode saber explicar por que escolheu ou rejeitou uma matéria prima, mas essa reconstrução acaba sendo parcial, pois ele não tem acesso a todos os motivos que realmente motivaram a escolha, ou seja, o explícito e o implícito se misturam (Sterelny, 2010, p. 293-4).

Alguns proponentes da GMU argumentam que as reflexões morais não estão conectadas com os julgamentos morais. Dwyer (2009), *e.g.*, aponta para as pesquisas de Haidt (2001) que demonstram o fenômeno do *moral dumbfounding*. Autores como Dwyer acreditam que esse fenômeno ocorre simplesmente porque os princípios que determinaram a decisão dessas pessoas

(*operative principles*) são inacessíveis à consciência (Mikhail, 2011, p. 83; Dupoux & Jacob, 2007, p. 374). Como vimos no capítulo anterior, as pesquisas de Haidt indicam que superestimamos a importância da reflexão consciente em nossos julgamentos morais. Em boa parte dos casos esse processo consciente serve apenas para encontrar uma racionalização para uma decisão intuitiva, *i.e.*, a nossa consciência desempenharia o importante papel social de encontrar justificáveis adequadas para os nossos julgamentos, mas não teria acesso às causas dos julgamentos intuitivos. No entanto, Haidt destaca que a reflexão não é apenas um epifenômeno da moral. Em algumas circunstâncias ela pode servir para modificar o julgamento intuitivo do próprio indivíduo e, em inúmeros casos, serve para alterar as intuições de outros membros do grupo, ou seja, exerce também influência social (Haidt & Bjorklund, 2008). Quando fazemos um julgamento moral não estamos guiando apenas nossa própria conduta, estamos envolvidos em uma atividade de valoração e orientação das condutas de outras pessoas. Esse é outro aspecto da moralidade que contrasta com o julgamento sintático. A persuasão moral não possui análogo na linguagem (Sterelny, 2012; Sterelny, 2010, p. 284).

Sterelny salienta que, além de podermos reconstruir boa parte dos princípios que motivam nossos julgamentos, podemos também modificar com o tempo as nossas intuições e emoções, por meio de processos conscientes de reflexão sobre as nossas convicções morais. No capítulo anterior já mencionei exemplos a respeito de como convicções morais oriundas de processos de raciocínio consciente podem afetar as nossas emoções e alterar os nossos julgamentos (nojo de vegetarianos em relação à carne; atitude em relação à homossexualidade; crueldade com animais; *etc.*). A mudança de convicção atrelada a uma adequação das práticas cotidianas à nova convicção acabam por alterar os tipos de objeto que provocam repugnância no indivíduo (Sterelny, 2010, p. 282-4 e 293-4).

Há, portanto, no âmbito da moralidade, uma relação diacrônica entre reações intuitivas e o pensamento consciente e articulado. Explicar essa interação entre a cognição tácita e a cognição explícita é um grande desafio enfrentado pela teoria da GMU, pois os seus defensores não dispõem de um modelo convincente. A analogia com a linguagem não pode auxiliar nessa tarefa, pois o processamento de normas sintáticas não depende de maneira alguma de raciocínio consciente (Sterelny, 2010, p. 282-4).

4.3 Desenvolvimento preestabelecido, aquisição precoce e facilidade de aprendizado

Os nativistas no campo da linguística defendem a sua posição indicando que a linguagem emerge nas crianças de forma previsível. Elas passam por estágios similares de desenvolvimento em idades semelhantes e atingem a competência linguística na mesma época. Entende-se que

isso é uma evidência de que a linguagem se desenvolve seguindo um cronograma endógeno (Prinz, 2008, p. 395). Nativistas morais também empregam esse tipo de argumento. Eles defendem a existência de um padrão de desenvolvimento da faculdade moral, mesmo perante grandes variações ambientais (Dwyer, Huebner & Hauser; 2010, p. 499-503). Harman, *e.g.*, considera improvável que seja necessário instrução explícita para as crianças adquirirem moralidade, embora a interação com outros membros da espécie seja necessária. Ele acredita ser possível que crianças que interajam somente entre si inventem sistemas morais, assim como crianças surdas criadas por pais com audição normal inventam linguagens de sinais (Harman, 2008, p. 348).

Prinz, entretanto, critica a ideia de que a moralidade se desenvolve segundo um cronograma preestabelecido através de uma análise do modelo mais famoso da teoria do desenvolvimento moral, desenvolvido por Lawrence Kohlberg³³. Segundo esse modelo, haveria seis níveis de desenvolvimento moral pelos quais nós passaríamos seguindo uma sequência fixa. Prinz aponta a existência de muitas falhas na teoria de Kohlberg. Em primeiro lugar, essa teoria trata do raciocínio moral, não do modo como formamos opiniões morais, isto é, ela ignora aquilo que apresentamos no primeiro capítulo: o raciocínio moral serve muitas vezes para justificar o julgamento, não para provocá-lo. Em segundo lugar, evidências empíricas não confirmaram uma progressão linear conforme os níveis estabelecidos por Kohlberg: as pessoas ora raciocinam conforme um nível mais avançado, ora regridem para um nível inferior, e a maioria dos adultos não passa sequer do quarto nível de desenvolvimento (Prinz, 2008, p. 395-7). Embora a crítica de Prinz seja bem fundamentada, ela se aplica apenas àqueles que seguem o modelo racionalista de Kohlberg.

Há outros modelos que escapam das críticas que Prinz dirige a Kohlberg. Haidt e Bjorklund, *e.g.*, ao tratarem do desenvolvimento da moralidade, empregam o conceito de externalização assistida. A moralidade seria melhor descrita como algo que emerge da criança de forma previsível, que é externalizado conforme um cronograma particular de desenvolvimento, ao invés de algo que é colocado na criança de fora para dentro, internalizado, de acordo com um cronograma estabelecido pela sociedade. O padrão de desenvolvimento repentino e similar em diversas culturas sugeriria a maturação de uma habilidade endógena, não o aprendizado de um conjunto de normas culturais (Haidt, 2001, p 826-7). Eles salientam, no entanto, que assim como ocorre com a linguagem e com o desenvolvimento sexual, a moralidade requer orientação e exemplos da cultura local para externalizar-se (Haidt & Bjorklund, 2008, p. 206).

³³ Para uma descrição detalhada do modelo de Kohlberg, ver Krebs, 2005.

Percepções envolvidas nos julgamentos morais emergem muito cedo. Nas primeiras horas de vida, os recém-nascidos choram ao escutar outras crianças chorarem. Isso poderia indicar que elas nascem prontas para experimentar uma forma rudimentar de empatia (Eisenberg *et al.*, 2003, p. 789). Antes de caminhar, crianças já são capazes de reconhecer e de valorar condutas como ajudar ou machucar terceiros. Hamlin, Wynn e Bloom (2007) apresentaram a crianças com idades entre 6 e 10 meses uma encenação com fantoches na qual um desses tentava subir uma encosta. Em algumas das apresentações, o fantoche era ajudado por outro que o empurrava para cima. Em outras, um terceiro fantoche aparecia em cima da encosta e golpeava o fantoche que subia, impedindo-o de chegar ao topo. Após essas apresentações, o fantoche que ajudava e o que atrapalhava eram postos em frente às crianças, e elas demonstravam uma forte preferência pelo primeiro. Os autores chegaram à seguinte conclusão sobre a experiência: “Our findings indicate that humans engage in social evaluation far earlier in development than previously thought, and support the view that the capacity to evaluate individuals on the basis of their social interactions is universal and unlearned” (Hamlin *et al.*, 2007, p. 559).

Esse tipo de experimento pode reforçar a ideia da existência de rudimentos de uma moralidade intuitiva em etapas iniciais do desenvolvimento de uma criança. Segundo Haidt e Kesebir, essa experiência pode indicar a existência de um sistema de percepção capaz de criar emoções positivas em relação a quem ajuda, e emoções negativas em relação a quem prejudica terceiros³⁴ (Haidt & Kesebir, 2010, p. 804).

O fato de algum traço aparecer bastante cedo no desenvolvimento favorece uma visão nativista, pois quanto antes um traço se manifestar menos tempo a cultura (*nurture*) terá para construí-lo. Bebês recém-nascidos com cerca de uma hora de vida, *e.g.*, podem imitar as expressões faciais de um adulto. Não parece plausível supor que essa capacidade seja uma construção cultural, sendo que até então o bebê passou a maior parte de sua vida junto ao peito de sua mãe (Hauser, 2007, p. 164).

Contudo, o fato de um indivíduo fazer algo muito cedo não significa necessariamente que não tenham adquirido esse conhecimento por observação (Prinz, 2012). Uma criança é capaz de associar uma palavra a um objeto a partir de uma única experiência; ela é capaz, por exemplo, de saber que aquele brinquedo em sua frente é um caminhão porque escutou alguma vez alguém chamá-lo daquela maneira.

³⁴ Há outras experiências similares acerca da existência de uma base inata para o comportamento social: ver Kinzler *et al.*, 2007; e Warneken & Tomasello, 2006.

Psicólogos que estudam o desenvolvimento cognitivo de crianças costumam elaborar experiências com bebês para tentar identificar possíveis conhecimentos em diferentes áreas como física, biologia e psicologia que sejam anteriores a qualquer forma de aprendizado. Muitas destas experiências adotam a premissa de que bebês fixam o olhar por mais tempo em situações inesperadas. Em uma famosa experiência feita com base no tempo que crianças olham para determinado evento, o psicólogo Renee Baillargeon (1987) concluiu que crianças com cerca de quatro meses já sabem identificar que um objeto continua a existir mesmo quando desaparece de seu campo de visão. Prinz destaca que esse tipo de experiência não demonstra que esse conhecimento não foi aprendido, pois desde que nascem as crianças vivenciam inúmeras experiências nas quais objetos que eles estão olhando desaparecem de sua visão e voltam a aparecer em algum momento posterior. De fato, situações como essa ocorrem cada vez que a criança fecha os olhos e, quando os reabre, o mundo ao seu redor continua inalterado (Prinz, 2012). De forma análoga, o desenvolvimento precoce da moralidade não implica necessariamente a existência de conhecimentos não aprendidos.

O argumento da pobreza de estímulos encontra suporte na aquisição da linguagem pelas crianças logo no começo da vida. Nesse curto período, elas não teriam tempo suficiente para escutar tal número de sentenças que pudesse ser suficiente para explicar o aprendizado. No caso da moral, os estímulos só parecem ser realmente pobres quando ela é concebida em termos de regras e princípios racionais — *e.g.* princípio do duplo efeito. Como já disse, as crianças recebem muito mais do que apenas instruções verbais. Os adultos reagem com mudanças em suas expressões às infrações cometidas pelas crianças (Sterelny, 2010). Quando essas reações são consideradas, é mais difícil sustentar a pobreza de estímulos. O que Mikhail chama de conhecimento moral precoce poderia ser explicado em termos de reforços emocionais. O desenvolvimento precoce pode ser explicado em termos de uma propensão emocional (*emotional preparedness*), sem a necessidade de apelarmos a princípios racionais prévios à experiência.

Na área da psicologia é universalmente aceito que algumas coisas são fáceis de aprender — ter medo de cobras —, enquanto outras são difíceis ou impossíveis — ter medo de flores ou odiar a justiça. Ninguém em psicologia estaria disposto a acreditar na teoria da *tabula rasa* (Haidt & Bjorklund, 2008, p. 183). Tooby, Cosmides e Barrett (2005) salientam como as crianças resistem aos esforços de seus pais para as fazerem valorizar ou desejar certas coisas. É extremamente difícil moldar uma criança quando o esforço é feito em sentido contrário ao que elas gostam naturalmente. Não exige muito trabalho, *e.g.*, fazer uma criança preferir balas a brócolis, a simpatia de outras crianças à aprovação de adultos, ou revidar uma agressão ao invés

de amar os inimigos. Produzir as preferências contrárias, por outro lado, será difícil. Haidt e Bjorklund salientam que esses vieses presentes no aprendizado explicam os fracassos das tentativas de construir sociedades utópicas:

The resistance of children to arbitrary or unusual socialization has been the downfall of many utopian efforts. Even if a charismatic leader can recruit a group of unusual adults able to believe in universal love while opposing all forms of hatred and jealousy, nobody has ever been able to raise the next generation of children to take on such unnatural beliefs (Haidt & Bjorklund, 2008, p. 201-2).

A ideia central desse tipo de modelo é que os seres humanos possuem fortes predisposições para desenvolver certas reações e preferências (Giroux, 2011, p. 292-3). A determinação daquilo que consideramos permissível, proibido ou obrigatório no âmbito moral pode ser resultado dessas propensões emocionais, sem que haja a necessidade da aplicação de um conjunto de princípios.

Mikhail, no entanto, acredita que só podemos descrever a competência moral de maneira adequada quando reconhecemos que ela envolve muitas das características de um código jurídico bastante desenvolvido, tais como teorias abstratas sobre crimes, condutas ilícitas, contratos e capacidade para agir. Ele acredita que, para fazermos julgamentos morais, empregamos uma série de capacidades cognitivas: precisamos de uma teoria da ação, para reconhecer que um comportamento particular é um exemplo de assassinato; e de uma teoria da mente para atribuir estados mentais, como crenças e desejos, a outras pessoas (Mikhail, 2008, p. 321; 2011, p. 104).

Quando compreendida dessa maneira, a competência moral não poderia, segundo Mikhail (2011), ser explicada apelando-se à instrução explícita ou a processos conhecidos de imitação, de internalização e de socialização. Ele ilustra essa afirmação com exemplos sobre o desenvolvimento de crianças. Com uma idade entre 3 e 4 anos elas já são capazes de empregar a intencionalidade como critério para distinguir duas ações que apresentam o mesmo resultado, mas têm diferentes motivações. Entre 4 e 5 anos já utilizam o princípio da proporcionalidade para determinar o nível apropriado de punição. Com 5 anos, demonstram entendimento sobre negligência e reparação. Entre 5 e 6 anos, consideram que falsas crenças sobre fatos podem inocentar, mas falsas crenças morais não. Com 8 anos, ao responderem aos *trolley problems*, admitem matar 1 pessoa para salvar 5, mas somente quando o meio escolhido não é errado, não há desproporção entre os efeitos negativos e positivos e não há melhor alternativa.³⁵ Em casos como esses, para conseguirmos explicar os resultados das observações, precisaríamos atribuir

³⁵ Esse tipo específico de caso refere-se à aplicação do princípio do duplo efeito e será analisado posteriormente.

conhecimento inconsciente e operações mentais complexas à criança que vão além daquilo que a ela foi ensinado.

As capacidades cognitivas mencionadas por Mikhail são essenciais para julgar moralmente uma ação, mas cabe destacar que elas desempenham diversas outras funções. As representações mentais que fazemos a respeito das crenças, desejos e objetivos fazem parte de diversos processos não atrelados à moralidade. A distinção entre ações intencionais e acidentais, *e.g.*, costuma ser relevante para os julgamentos morais, pois muitas ações só são consideradas condenáveis quando praticadas intencionalmente. No entanto, ao observarmos uma criança brincar com uma bola, somos capazes de distinguir se a bola escapou das mãos dela ou se quicou a bola de propósito, embora isso não tenha qualquer relevância moral. De modo semelhante, quando enxergamos uma criança cortando um limão, nós atribuímos um objetivo àquele ato, *e.g.*, preparar uma limonada. Nós inferimos constantemente propriedades invisíveis da mente a partir de verificações indiretas, tais como o local para onde outra pessoa está olhando, o objeto que ela está tentando alcançar ou onde ela esteve. Teorias da mente e da ação são necessárias para compreender a intenção e a ação dessa criança, mas não estão vinculadas a um julgamento moral. Essas capacidades seriam como as capacidades para percepção visual ou para a memória. Elas podem ser utilizadas no julgamento moral, mas não são específicas desse domínio, e podem ser utilizadas para um conjunto de outras funções (Sripada, 2008, p. 362; Hauser, 2007, p. 50-1).

Esse tipo de advertência, contudo, é relevante apenas para a refutação de modelos nativistas que adotam uma concepção evolutiva de ‘inato.’ Jesse Prinz adota esse tipo de concepção e dedica boa parte de seus trabalhos a demonstrar que as capacidades envolvidas nos julgamentos morais não são específicas a esse domínio. Abordar o problema dessa forma não nos ajuda a compreender as origens da moralidade, o modo como adquirimos essa capacidade e a influência exercida pelas variantes culturais. Afirmar, *e.g.*, que o fato de experimentarmos aflição e estresse ao presenciarmos o sofrimento de outra pessoa (*vicarious distress*) não é uma disposição exclusivamente moral, não responde se essa disposição é parte da explicação do fato de sociedades humanas terem normas morais (Giroux, 2011, p. 287). Contudo, é importante frisar, a demonstração de que as capacidades cognitivas envolvidas nos julgamentos morais não são específicas a esse domínio constitui um entrave a teorias como a da gramática moral que postulam a existência de um domínio específico responsável pelos julgamentos morais.

4.4 Universalidade das normas morais

A moralidade existe em todas as sociedades humanas de que temos conhecimento e em praticamente todo indivíduo, desenvolve-se mesmo sem instrução formal e sem esforço deliberado (Joyce, 2006; Ayala, 2010, p. 9016). Defensores de modelos nativistas apontam para estudos antropológicos que demonstram a existência de normas e valores similares em diversas culturas. Chandra Sekhar Sripada lista os seguintes exemplos:

[M]ost societies have rules that prohibit killing and physical assault (Brown, 1991). Also, most societies have rules promoting sharing, reciprocating, and helping, at least under some circumstances (Cashdan, 1989). Most societies have rules regulating sexual behavior among various members of society, and especially among adolescents (Bourguignon & Greenbaum, 1973). And most societies have at least some rules that promote egalitarianism and social equality. For example, in nearly all hunter-gatherer groups, attempts by individuals to reap a disproportionate share of resources, women, or power are disapproved of sharply (Boehm, 1999). Examples like these could be multiplied easily in domains such as social justice, kinship, marriage, and many others (2008, p. 322)."

Se essas normas fossem apenas construções culturais sem qualquer raiz na condição humana, como poderíamos explicar essas similaridades?

A antiguidade e universalidade das normas, segundo autores nativistas, seria uma evidência de que a cognição normativa é produto da evolução. Quando uma característica é antiga e universal, haveria duas possibilidades. Ou ela poderia ser aprendida facilmente por aprendizado social ou individual — *e. g.*, a crença de que o sol nasce toda manhã —; ou haveria um sistema de desenvolvimento que garantiria o seu desenvolvimento regular. Quando se trata dessa segunda hipótese, a antiguidade e universalidade serviriam como indício de seleção natural. Se não conseguirmos explicar como alguém adquire a capacidade para ter atitudes normativas em relação a certos comportamentos, é plausível que a cognição normativa tenha evoluído (Machery & Mallon, 2010).

Defensores de modelos construtivistas empregam dois tipos de estratégia para rejeitar a universalidade da moral. A primeira delas é buscar exceções às supostas intuições ou normas universais (Giroux, 2011, p. 284). Contra a proibição de assassinato, *e.g.*, eles indicam o ritual de passagem para a vida adulta dos Llongot, uma tribo Filipina na qual os garotos devem trazer a cabeça de uma pessoa inocente de uma tribo vizinha (Rosaldo, p. 137-176). Contra a atribuição de uma especial dignidade ou status aos seres humanos, há inúmeros registros de canibalismo ao longo da história. Quanto à intuição de que danos causados por ação são mais graves, do ponto de vista moral, do que os causados por omissão, os Mayans constituem uma exceção (Abarbanell & Hauser, 2010, p. 207).³⁶ A segunda estratégia, uma vez enfraquecida a

³⁶ Segundo Sterelny, nós julgamos que as omissões são menos graves em razão de elas de afastarem dos paradigmas de condutas puníveis. Sterelny sugere que os Mayans possibilitam o teste de seu modelo de

universalidade, é oferecer uma explicação não nativista para a existência de normas e valores quase-universais. Essas explicações teriam a vantagem de ser mais simples ou parcimoniosas do que as alternativas nativistas, principalmente porque elas não dependem de faculdades específicas ou de módulos especializados, cuja existência é controversa (Giroux, 2011, p. 284).

Construtivistas procuram demonstrar que embora a moralidade exista, à primeira vista, em todas as culturas, a variação nas regras morais é bastante grande. Assim, o modelo de princípios e parâmetros oriundo da analogia linguística, em razão de basear-se em um grupo de parâmetros discretos e relativamente rígidos, não acomodaria essa grande variação no conteúdo das normas morais presentes nos diferentes grupos (Sripada, 2008). Conforme Chomsky, a amplitude de variação das línguas entre os grupos humanos é restringida por parâmetros universais que admitem apenas certos tipos restritos de variação. A existência desses parâmetros torna possível fazer generalizações sem qualquer exceção sobre a estrutura das línguas humanas e torna possível prever quais os tipos de línguas humanas são possíveis ou impossíveis. Se a moral estivesse vinculada a esse tipo de parâmetros, esperaríamos encontrar também generalizações sem qualquer exceção. Susan Dwyer, defensora da analogia com a linguagem, acredita que princípios morais universais restringem o espaço das possíveis regras. Ela sugere que nenhum ser moral “normal” poderia criar a regra de que é bom torturar bebês por diversão (Dwyer, 2008, p. 414).

Dados etnográficos sugerem que o padrão de variação de normas é muito mais complexo, sutil e variado do que sugere o modelo parametrizado. Em relação, por exemplo, às normas contra condutas lesivas, muitos grupos permitem que algumas dessas condutas sejam cometidas contra membros do próprio grupo, pelo menos em algumas circunstâncias, e os grupos divergem de maneira complicada em relação ao tipo de dano ou lesão que é permitido (sacrifício, perfurações, circuncisões, flagelação, *etc.*). Em muitas sociedades, a violência contra certos subgrupos — mulheres, crianças, castas marginalizadas — é permitida (LeVine & Campbell, 1972; Edgerton, 1992; Sripada, 2008). Há, ainda, registros de normas claramente mal adaptativas, especialmente em comunidades utópicas e em ordens religiosas, como a dos Shakers (Sripada, 2008, p. 329-332).

Prinz defende que a amplitude de variação das normas morais é praticamente ilimitada. Ao contrário do que ocorre com a linguagem, na qual certas gramáticas nunca ocorrem, a variação moral parece ser aberta. Respondendo especificamente ao exemplo de Dwyer, ele acredita ser possível ensinar um povo a torturar bebês como forma de recreação. Para defender

reconhecimento de padrões. Se ele estiver correto, o conjunto de exemplos relacionados à moralidade ao qual as crianças Mayans são expostas deve ser bastante atípico. A experiência narrativa das crianças dessa tribo deve estar repleta de exemplos desprezíveis de danos causados pela falta de ação (Sterelny, 2010, p. 289-90).

essa hipótese, ele menciona os esportes sangrentos praticados pelos romanos, o canibalismo dos astecas, as brigas de galo, os sacrifícios de crianças pelos fenícios, o infanticídio praticado sistematicamente pelos gregos. Enfim, segundo Prinz, a crueldade seria um traço comum da existência humana que possibilita facilmente o surgimento da tortura (Prinz, 2008).

Por outro lado, nativistas como Mikhail consideram essas afirmações sobre a variabilidade da moral insustentáveis. Segundo ele, a competência moral é muito mais restrita do que a competência linguística. Os sistemas linguísticos adquiridos pelas crianças não são apenas discrepantes, mas mutuamente ininteligíveis. Uma criança que cresce falando Inglês é tipicamente incapaz de identificar a fronteira entre palavras do Japonês e do Árabe, ou de dizer se construções nessas línguas são gramaticais ou não. No domínio da moral não ocorre nada semelhante. Se, *e.g.*, um homem atira em uma pessoa quando acreditava tratar-se de um tronco de árvore, as intuições despertadas são semelhantes em pessoas de diferentes culturas.

O argumento de Mikhail, contudo, não comprova que a variação presente na linguagem é maior. O fato de falantes de línguas distintas não se compreenderem mutuamente não tem relação direta com as variações gramaticais subjacentes. Duas línguas poderiam ser iguais sob o ponto de vista gramatical, mas adotar significantes distintos para expressar os mesmos significados. A ininteligibilidade está associada ao fato de que a linguagem envolve a utilização de códigos bastante complexos. Quem não conhece o código não pode compreender a língua. Imaginemos, *e.g.*, um falante analfabeto de Português. Se nós lermos para ele em voz alta uma frase qualquer — João foi ao mercado — ele a compreenderá, mas se apresentarmos para ele a mesma frase escrita em uma folha de papel ele será incapaz de compreender, ou seja, a mesma língua torna-se ininteligível apenas por estar codificada de maneira distinta.

Além desse argumento, Mikhail tenta resgatar a clássica ideia racionalista de que certas noções jurídicas não seriam fruto de aprendizado. Distinções tipicamente jurídicas baseadas em conceitos de causação, intenção e comportamento voluntário seriam universais. Em todas as línguas naturais haveria palavras para expressar conceitos deônticos básicos — *may*, *must*, *must not*. Conceitos como esse compreenderiam o esquema de categorização básico dos sistemas normativos humanos (Mikhail, 2011, p. 101-104). Essas noções explicariam por que os indivíduos frequentemente concordam sobre como analisar os principais componentes das ações humanas: ato, intenção, motivo, causa e efeito, consequências próximas e remotas, *etc.* (Mikhail, 2008, p. 356-7; Mikhail, 2011, p. 57).

Como já vimos, essas noções mencionadas por Mikhail, que fazem parte da análise estrutural de uma ação (causas, consequências, intenções, *etc.*), extrapolam o campo da moral. Elas integram a análise de qualquer tipo de ação, por mais estranha que ela seja ao campo da

moralidade. Portanto, esse tipo de argumento não serve para sustentar a existência de uma faculdade moral específica, como aquela defendida pelos teóricos da GMU.

Quanto à alegação de que existem conceitos deônticos básicos em todas as línguas naturais, ela se confunde com a própria alegação de que a moralidade se manifesta em todas as culturas, pois é difícil imaginar de que maneira algum grupo humano poderia valorar condutas moralmente sem recorrer a conceitos deônticos. Partindo do pressuposto de que todas as comunidades humanas conhecidas possuem linguagem, como exatamente seria pensável que os integrantes de um grupo qualquer pudessem julgar moralmente sem desenvolver conceitos deônticos? Julgar moralmente uma conduta envolve julgar se ela é obrigatória, proibida ou permitida. Sem esses conceitos não há como falar em julgamento moral. Portanto, o fato de as línguas naturais apresentarem esses conceitos se equipara à evidência antropológica de que todas as culturas conhecidas apresentam alguma forma de moralidade.

As grandes variações e as supostas exceções apresentadas por si só não são suficientes para refutar a pretensão de universalidade da moralidade — assim como o registro antropológico não pode por si só provar o universalismo ou o nativismo moral em suas diversas vertentes —, uma vez que as exceções podem ser acomodadas pelos nativistas (Giroux, 2011, p. 284). Em resposta aos contraexemplos, eles fazem, *e.g.*, uma distinção entre universalismo fundacional (*foundational*) e universalismo de conteúdo (*content*), defendendo que os contraexemplos refutam apenas o segundo tipo, na medida em que demonstram que o conteúdo das normas varia de forma significativa de uma sociedade para a outra. Contudo, eles não demonstram a inexistência de uma similaridade nas bases que sustentam a formação da moral (*foundational similarity*). As moralidades mais distintas podem repousar sobre os mesmos princípios ou disposições básicas, mas apresentar normas muito diferentes em razão da maneira como essas disposições são combinadas (Giroux, 2013, p. 173).

Outro tipo de objeção à universalidade da moral se baseia na distinção entre normas morais e outros tipos de normas. Machery e Mallon (2010), por exemplo, defendem que os estudos antropológicos utilizados para sustentar que a moralidade está presente em todas as sociedades humanas conhecidas na verdade comprovam apenas que há alguma forma de norma em todas as sociedades, mas não comprovam que em todas essas sociedades as normas presentes são compreendidas como normas morais. Eles chamam atenção para o fato de que há diversas formas de julgamento normativo, e o julgamento moral é apenas uma dessas formas. Nas sociedades ocidentais modernas essa diversidade de tipos de normas é marcante. Nós temos normas relacionadas à racionalidade (como devemos pensar), à estética, às convenções sociais, à prudência (se você quiser viver bastante, deve parar de fumar), à religião, ao direito, *etc.*

Nas sociedades não ocidentais e nas sociedades tradicionais não há essa diversidade tão grande a respeito de como as normas são compreendidas, e a moralidade costuma ser muito mais abrangente. Essas sociedades moralizam práticas que não têm qualquer relação com a produção de dor ou de danos a algum indivíduo (Haidt, 2012). Embora de fato muitos estudos falhem ao não distinguir entre normas morais e outros tipos de normas, os indícios que temos apontam no sentido de que as sociedades modernas ocidentais são aquelas nas quais a moralidade está menos presente. Nas outras sociedades, a moralidade não apenas está presente, como ocupa posição muito mais relevante.

Apesar da abrangência das normas morais variar bastante, algumas de suas características são comuns. Nos mais diversos grupos humanos há um conjunto de normas que regulam as condutas dos indivíduos e que apresentam as características elencadas no segundo capítulo. Quando as normas morais são caracterizadas dessa forma — como predominantes, válidas independentemente da sanção de alguma autoridade específica, extensíveis aos demais membros da comunidade relevante, inescapáveis, destinadas a frear comportamentos em nome de algo considerado importante por outros indivíduos — não deve haver controvérsia sobre a sua quase-universalidade. O problema com essa caracterização é que ela pode abranger certos tipos de normas que não são consideradas morais nas sociedades ocidentais modernas, tais como as normas de etiqueta (Foot, 1972). No entanto, isso acaba sendo o resultado da concepção estreita e peculiar de moralidade que vigora nas sociedades modernas ocidentais.

Cabe também destacar que a universalidade de uma característica não implica necessariamente que ela seja uma adaptação ou que não seja resultado de processos de aprendizado, basta pensar em coisas como o uso do fogo, armas, vestimentas, casamento, *etc.* Alguns dos desafios enfrentados pelos seres humanos são iguais em todas as partes do mundo e eles têm recursos cognitivos semelhantes para solucioná-los; desta maneira é natural que as mesmas soluções para esses desafios apareçam. Prinz acredita que a necessidade universal de obter estabilidade social garante que algum sistema de normas morais seja desenvolvido (Prinz, 2008, p. 372).

Ele atribui a presença universal da moralidade a dois traços humanos que não seriam propriamente morais. Em primeiro lugar, nossa vida é rica em emoções. Prinz adota uma teoria sentimentalista a respeito da teoria da moralidade, ou seja, ele compreende que os julgamentos morais são baseados em emoções, como raiva, nojo, culpa e vergonha. O segundo fator seria a nossa capacidade para imitação. Os seres humanos são grandes imitadores. Outros primatas também podem aprender por imitação, mas com muito menos eficiência e precisão. No âmbito da moral, seria relevante a capacidade que temos de imitar as atitudes dos nossos pais. Quando

punimos um cachorro, ele pode aprender a obedecer, mas não aprenderá a punir outros cachorros que não obedecerem. Com os outros primatas provavelmente ocorreria a mesma coisa. No entanto, quando um pai pune uma filha, em questão de pouco tempo, ela estará punindo o irmão mais novo pela prática do mesmo ato. Em todas as culturas, os pais ficam irritados com os seus filhos, o que provoca emoções de culpa nas crianças e, após muitas repetições, faz com que as próprias crianças tendam a ficar irritadas com aquele tipo de ato que inicialmente irritou os seus pais. Assim, surgiriam regras que se espalhariam pela família e poderiam ser calibradas e coordenadas em âmbitos maiores, como a vila na qual a família está inserida, conduzindo à conformidade no grupo e à geração de um sistema de normas. As emoções em conjunto com a capacidade humana de aprendizado social tornariam praticamente inevitável o surgimento da moralidade (Prinz, 2012).

A existência de uma solução universal para um problema universal de coordenar condutas explicaria a quase-universalidade de algumas normas morais (Giroux, 2011, p. 288). As pessoas reconhecem a possibilidade de moldar comportamentos condicionando respostas emocionais e a utilizam para resolver o problema de coordenação. Ao redor do mundo, os pais penalizam os seus filhos a fim de fazê-los comportarem-se da maneira desejada, e essas penalidades desencadeiam consequências emocionais. Esse processo de educação poderia ser um primeiro passo para explicar o surgimento das regras morais — *the building blocks of morality* (Prinz, 2008, p. 386-7).

No entanto, há alguns problemas com a tese de Prinz. Em primeiro lugar, as pressões para que os indivíduos coordenem as suas ações não parece suficiente para explicar a universalidade de alguns princípios morais básicos. Não está claro de que forma apenas essa necessidade de coordenação poderia explicar a extensão das similaridades morais em diferentes culturas. Além disso, se não fôssemos receptivos a normas que promovem a cooperação, poderíamos simplesmente não chegar a uma solução para esses problemas de coordenação. Se todos nós tivéssemos uma inclinação para enganar, trair e agir como *free-riders*, normas que promovem a cooperação não conseguiriam se impor e as sociedades humanas poderiam simplesmente desaparecer. Giroux propõe o exemplo de uma sociedade composta apenas por psicopatas. Tendo em vista a ausência de empatia em pessoas desse tipo, seria difícil imaginar que normas que promovem a cooperação seriam aceitas e aplicadas. O mais provável é que a sociedade entrasse em colapso, resultando em um estado de natureza hobbesiano (Giroux, 2013, p. 181-2). Há bons motivos para suspeitar que as disposições humanas resultantes de um longo

processo de seleção natural não sejam tão fracas quanto Prinz supõe³⁷ ou que mecanismos de condicionamento sejam forças tão irresistíveis (Giroux, 2013, p. 176).

De fato, o processo de educação descrito por Prinz não explica completamente o surgimento dos julgamentos morais. Ele reconhece a existência de uma base biológica emocional que faz parte da moralidade e a torna possível, mas ele se recusa a dizer que a moralidade é ‘inata’ por compreender que ela não envolve a utilização de mecanismos mentais específicos (*mental machinery*). Devemos lembrar que ele considera inato apenas aquilo que constitui uma adaptação no sentido biológico do termo. Desta forma, embora ele se posicione como um antinativista, poderíamos afirmar que até mesmo Prinz reconhece que grande parte daquilo que constitui e influencia a moralidade é produto da evolução biológica.

Uma vez constatado que a moralidade se desenvolve nos diversos grupos humanos existentes e que esses grupos vivem em ambientes muito distintos, podemos concluir que o desenvolvimento da moralidade é de alguma maneira canalizado. No entanto, não é tão simples identificar de qual tipo de canalização se trata. Teorias nativistas, como a da GMU, costumam defender que essa canalização resulta de fatores genéticos. Por outro lado, filósofos que adotam uma posição construtivista em relação à moralidade enfatizam que a canalização pode também decorrer de outros aspectos. Esse é o caso de Sterelny, *e.g.*, que enfatiza a *construção de nichos*. A teoria dos domínios morais, que veremos a seguir, e a análise que faremos no último capítulo sobre os modelos epidemiológicos de evolução cultural poderão nos ajudar a decidir qual dessas duas hipóteses é mais plausível.

Em suma, há uma série de possíveis explicações para a universalidade de uma determinada característica, de forma que a simples constatação da universalidade da moral não implica, por si só, a adoção de alguma explicação adaptacionista, tal como a oferecida pela GMU.

³⁷ Ver, nesse sentido, os exemplos mencionados anteriormente acerca de como é difícil educar uma criança para fazer certas coisas.

5 Teoria dos domínios morais: Haidt

Jonathan Haidt analisou listas de virtudes existentes ao redor do mundo e percebeu algumas semelhanças e superposições. Em diferentes culturas, os indivíduos apresentam alguns dos mesmos tipos de reações automáticas e emocionais às interações sociais (Ekman *et al.*, 1969). Com base nisso, psicólogos e antropólogos culturais já haviam proposto que os sistemas morais são baseados em um pequeno conjunto de valores morais universais (*e.g.*, Rozin *et al.*, 1999; Shweder *et al.* 1997).

Haidt começou a se perguntar se não haveria, no âmbito da moral, algo semelhante aos receptores que temos para perceber os sabores. Ele ilustra essa analogia da seguinte forma: na grande maioria das culturas existem um ou mais tipos de bebidas doces que são largamente consumidas, normalmente derivadas de frutas locais ou, nos países industrializados, de açúcar e aromatizantes. Seria absurdo imaginar que temos diferentes receptores para suco de manga, suco de maçã e Coca-Cola. Há no caso em questão apenas um tipo de receptor capaz de identificar sabores adocicados, mas cada cultura inventa diferentes formas de acioná-lo.

No âmbito da moral, apesar da variação cultural de normas e práticas, haveria um pequeno conjunto de intuições morais facilmente encontráveis em todas as sociedades, mas que poderiam ser desencadeadas de diversas formas, conforme as variações culturais. As nossas mentes têm o potencial para incluir as mais diversas práticas nos domínios da moral. No entanto, apenas algumas dessas práticas, a partir de nossas experiências culturais, serão incluídas, ao longo de nossa infância, naquilo que compreendemos por moralidade. Outras tantas práticas permanecerão desconectadas daquilo que forma a rede de valores e sentidos compartilhados que compõe a nossa matriz moral (Haidt, 2012, cap. 5).

As cozinhas são produtos da cultura e cada uma é única e possui um conjunto de ingredientes principais. No entanto, elas são construídas a partir de um sistema sensorial que inclui apenas 5 tipos de receptores para os gostos e de um sistema olfativo mais complexo. Os tipos de papilas gustativas trazem óbvios benefícios adaptativos: o sabor doce é útil para reconhecer frutas e indicar que o alimento é seguro; o sabor amargo indica a presença de toxinas e alerta para o perigo; o receptor de glutamato indica a presença de carne e assim por diante. A estrutura da nossa língua, de nosso nariz e do nosso cérebro estabelece restrições aos tipos possíveis de cozinha, mas deixa bastante espaço aberto à criatividade. A ideia de Haidt é que haveria algo semelhante no domínio da moral, ou seja, haveria um pequeno conjunto de intuições morais sob a grande diversidade de sistemas de normas morais (Haidt & Bjorklund, 2008, p. 201-2). Tanto a moralidade quanto as diferentes cozinhas seriam construções culturais,

influenciadas por diversos fatores históricos e ambientais, mas não seriam completamente flexíveis. Não podemos ter um tipo de cozinha baseada em cascas de árvores ou predominantemente em sabores amargos, pois, para que uma culinária se estabeleça, ela precisa agradar línguas equipadas com uma variedade limitada de papilas gustativas. De modo semelhante, uma matriz moral precisa ser compatível com mentes equipadas com determinados tipos de receptores sociais (Haidt, 2012, cap. 6).

Haidt e Joseph (2004) procuraram identificar, a partir de 5 pesquisas sobre características universais (Brown, 1991; Fiske, 1992; Schwartz & Bilsky, 1990; Shweder *et al.*, 1997; de Waal, 1996), princípios comuns da moralidade humana. Após tentarem listar todas as coisas que humanos e chimpanzés pareciam valorar no comportamento dos outros, eles as dividiram em cinco categorias: (a) sensibilidade ou aversão a sinais de dor e sofrimento em outros (*harm/care*); (b) respostas negativas àqueles que falham em retribuir favores (*fairness/reciprocity*); (c) raiva contra aqueles que falham em demonstrar sinais de deferência e respeito (*authority/respect*); (d) emoções relacionadas a nojo e repulsa, necessárias para explicar regras morais sobre comida, sexo, menstruação e manuseio de cadáveres (*purity/sanctity*); (e) atitudes em relação às fronteiras entre os grupos (*in-group/out-group*)³⁸. Esses cinco grupos de intuições seriam as bases da moralidade (Haidt & Bjorklund, 2008, p. 202-203) e cada um deles representaria um módulo mental e se relacionaria a diferentes famílias de emoções. O sofrimento provoca empatia e compaixão; desrespeito à hierarquia provoca ressentimento; violação da reciprocidade provoca raiva e culpa; violações relacionadas ao domínio da pureza provocam nojo. Esses domínios seriam universais, mas cada cultura poderia determinar os conteúdos específicos de cada um deles (Haidt & Joseph, 2004).

5.1 As origens dos cinco domínios morais

Ao elaborar a sua teoria, Haidt procura identificar as conexões entre as virtudes existentes e as teorias evolutivas bem estabelecidas (Haidt, 2012, cap. 6). Ele busca identificar os desafios adaptativos da vida social de nossos ancestrais e conectá-los às virtudes presentes de alguma forma em diversas culturas. Para cada uma das emoções e intuições associadas a essas virtudes, ele traça uma história evolutiva que procura explicar quais eram as circunstâncias que as desencadeavam e, assim, estabelecer uma ligação entre as observações antropológicas e a teoria evolutiva, embora reconheça que as explicações por ele propostas para a origem das emoções são apenas “educated guesses.”

³⁸ Em alguns outros trabalhos Haidt designa esse domínio “Loyalty/Betrayal Foundation” (Haidt, 2012)

Como vimos anteriormente, ao tratarmos da distinção entre normas morais e convencionais, muitos pesquisadores³⁹ atribuem àquilo que Haidt chama de *care/harm foundation* o papel central na construção da moralidade. A origem desse domínio, segundo Haidt, estaria relacionada ao modo como os mamíferos se reproduzem e cuidam de suas crias. O próprio ato de amamentar torna a maternidade muito mais custosa. Como consequência, os mamíferos são obrigados a concentrar os seus recursos em um número bastante limitado de filhos. Eles investem muito mais em cada uma de suas crias e enfrentam o desafio de tomar conta delas por um longo tempo. No decorrer do processo evolutivo, aquelas mães com maior sensibilidade aos sinais de sofrimento de suas crias tinham maiores chances de conseguir que seus filhos chegassem até a idade reprodutiva e, portanto, tinham maiores chances de propagar os seus próprios genes. Ou seja, a evolução pode ter favorecido as fêmeas e, em menor medida, os machos que tinham reações automáticas aos sinais de sofrimento e necessidade⁴⁰. Posteriormente, por uma série de processos culturais e/ou biológicos, as reações ao sofrimento e à necessidade passaram a ser desencadeadas em várias outras situações, não apenas nas relações intrafamiliares (Haidt, 2012, cap. 7)

Por sua vez, a origem do domínio da justiça e da reciprocidade poderia ser explicada pelo mecanismo de altruísmo recíproco proposto por Trivers para explicar a cooperação⁴¹. Trivers observou que o altruísmo poderia ser explicado pela evolução desde que os indivíduos pudessem lembrar de suas interações prévias e limitar a sua cooperação a indivíduos que provavelmente retribuiriam os favores prestados. Caso contrário, indivíduos que não cooperam teriam uma vantagem adaptativa em relação aos demais: usufruiriam dos benefícios da cooperação sem precisarem arcar com quaisquer custos. Por milhões de anos, nossos ancestrais enfrentaram o desafio de colher os benefícios da cooperação sem serem explorados por aproveitadores. Modelos matemáticos demonstram que em cenários como esses a melhor estratégia é aquela chamada ‘*tit for tat*’ (Axelrod, 1984), ou seja, cooperar na primeira interação e reproduzir nas interações futuras o comportamento que a outra parte praticou na última interação: cooperar se ela cooperou; desertar se ela desertou. Aqueles indivíduos cujas emoções os motivavam a agir dessa forma teriam uma vantagem evolutiva em relação aos demais. Desta forma, os mecanismos iniciais que desencadeavam reações emocionais associadas a esse domínio eram atos de cooperação e de egoísmo que outros indivíduos praticavam em relação a nós. Quando nos deparávamos com um ato egoísta, sentíamos raiva e desprezo por aquele que

³⁹ O filósofo Shaun Nichols, *e.g.*, adota essa posição (2008, p. 266).

⁴⁰ Hipótese baseada na teoria do apego.

⁴¹ Para uma análise detalhada sobre diferentes explicações para a evolução da cooperação na espécie humana, ver Abrantes (2014a); Almeida e Abrantes (2012); e Almeida (2011).

tentou nos enganar ou tirar alguma vantagem (Rozin *et al.* 1999; Haidt, 2012, cap. 7). Os grupos de hominíneos enfrentaram pressões seletivas que levaram ao surgimento de sociedades igualitárias de caçadores e coletores nas quais havia grande cooperação. Dentro desse contexto, aqueles que possuíam reações instintivas adequadas às interações sociais tinham vantagem em relação aos demais. As emoções associadas ao domínio da justiça e da reciprocidade teriam desempenhado um papel fundamental na regulação desse tipo de conduta.

A universalidade desse domínio é confirmada por estudos que aplicaram o jogo do Ultimato em vários tipos diferentes de sociedades, inclusive sociedades de coletores, de pastores nômades, de horticultores que empregam a coivara, de pequenos agricultores *etc.* (Henrich *et al.*, 2001; Henrich *et al.*, 2005). Nesse jogo, um dos jogadores recebe uma quantia de dinheiro e deve propor a forma de dividi-lo com um segundo jogador. Se esse aceitar a proposta, cada um recebe a parte acordada; se rejeitar, ninguém recebe nada. As ofertas propostas variaram muito entre as diferentes culturas — normalmente de 15 a 50% do montante total. O que era considerado justo (*fair*) ou iníquo dependia das restrições presentes em cada cultura, mas, de maneira geral, em cada sociedade estava presente um senso de justiça (*fairness*) e os indivíduos estavam dispostos a arcar com custos, *i.e.*, abrir mão da parte do dinheiro que lhes caberia, para punir os que praticavam injustiças (Hauser, 2007, p. 83-4). Caso os indivíduos agissem com base simplesmente em uma racionalidade econômica, não seria possível explicar esses resultados.

No que diz respeito às emoções e aos instintos associados ao domínio da autoridade e respeito, eles teriam surgido em sociedades hierarquizadas (Haidt, 2012, cap. 7). Até hoje as espécies mais próximas do *Homo sapiens*, como os chimpanzés e gorilas, vivem em grupos organizados desta maneira. O grande dimorfismo sexual presente em espécies hominíneas anteriores à nossa também indica que o igualitarismo é um fenômeno relativamente recente, e, mesmo em grupos igualitários, a relação entre pais e filhos é baseada na hierarquia. Além do mais, a estrutura hierarquizada também é a regra nas sociedades humanas do Holoceno. Como destaca o antropólogo Alan Fiske, as relações presentes nesse tipo de sociedade não se estabelecem apenas com base no poder de coerção; tampouco são necessariamente explorativas em relação aos que ocupam posições inferiores, elas dependem de uma percepção de legitimidade das assimetrias. Nesse tipo de organização cabe aos subordinados obedecer, respeitar e demonstrar deferência, enquanto os superiores têm responsabilidades em relação à proteção dos subordinados e aplicação das normas. Há inúmeros exemplos de organizações desse tipo, como as hierarquias militares e religiosas, além da própria forma de organização presente nas famílias (Fiske, 2016).

Nesse contexto, comportamentos adequados à ordem hierárquica teriam sido favorecidos. Os mecanismos desencadeadores desses módulos seriam padrões de comportamento e de aparência que distinguiriam as classes superiores e inferiores. Assim como os chimpanzés, as pessoas têm a capacidade de recordar qual a posição dos outros membros do grupo. Quando alguém nega ou subverte essa ordem, sentimos determinadas emoções que nos motivam a defender a estrutura social, ainda que não sejamos afetados diretamente (Haidt, 2012, cap. 7).

Há experiências que demonstram como a obediência à autoridade desempenha um papel relevante no comportamento humano. No começo dos anos 60, o psicólogo Stanley Milgram conduziu um estudo clássico sobre o tema (Milgram, 1974). O estudo simulava uma situação de aprendizado na qual o indivíduo que desempenhava a função de professor era orientado pelo cientista a dar choques cada vez mais intensos sempre que o aprendiz errasse uma questão. Supostamente o estudo visava aferir se a punição, na forma de choques, melhorava o aprendizado. Na verdade, o objetivo real de Milgram era testar a obediência dos participantes. Ele verificou que grande parte dos participantes estava disposta a infligir choques em níveis extremamente altos, mesmo diante de apelos no sentido contrário por parte do suposto aprendiz. Embora eles acreditassem ser errado causar sofrimento e danos físicos a outras pessoas, eles estavam dispostos a ignorar essa proibição em obediência a uma autoridade que não dispunha de qualquer poder real. Milgram chegou à seguinte conclusão sobre o experimento:

Subjects have learned from childhood that it is a fundamental breach of moral conduct to hurt another person against his will. Yet, almost half of the subjects abandon this tenet in following the instructions of an authority who has no special powers to enforce his commands. To disobey would bring no material loss or punishment. It is clear from the remarks and behavior of many participants that in punishing the victim they were often acting against their own values. Subjects often expressed disapproval of shocking a man in the face of his objections, and others denounced it as stupid and senseless. Yet many followed the experimental commands (1974, p.41).

A experiência de Milgram foi replicada em diversos lugares do mundo e sugere que a obediência à autoridade é universal, embora o grau de obediência varie consideravelmente em diferentes culturas. Em todos os países, os participantes estavam dispostos a aplicar choques em níveis elevados, mas havia variação relevante entre eles. Na Alemanha, 85% dos participantes estavam dispostos a aplicar o nível máximo de choque (450 volts), enquanto na Austrália o percentual caía para 40% (Hauser, 2007, p. 140-1). A cultura pode alterar os níveis de obediência de seus membros, mas a capacidade para comandar e obedecer seria uma capacidade mental resultante da evolução e compartilhada com outros primatas (Hauser, 2007, p. 140-1).

Com relação ao domínio da pureza e da santidade, ele teria a sua origem no chamado dilema do onívoro. Onívoros têm a necessidade de procurar novos alimentos, mas devem manter

cautela em relação a eles até terem certeza de que são seguros. A emoção de nojo teria surgido como uma forma de otimizar as respostas a esse dilema. Além disso, quando os hominíneos abandonaram as árvores e passaram a viver em grupos maiores, eles aumentaram os riscos de contrair infecções uns dos outros e dos dejetos produzidos. A emoção de nojo faria parte daquilo que o psicólogo Mark Schaller chama sistema imunológico comportamental, *i.e.*, de um conjunto de módulos desencadeados por sinais de infecção ou doenças em outras pessoas e destinados a nos afastar dessas pessoas (Schaller & Park, 2011). Portanto, inicialmente os mecanismos desencadeadores da emoção de nojo relacionavam-se à necessidade de evitar contato com parasitas e agentes patogênicos. Essa emoção seria a base do domínio da pureza e da santidade, que nos leva a considerar certas coisas como intocáveis, seja porque são ‘poluídas’ ou sagradas (Haidt, 2012).

A presença desse domínio nas modernas sociedades ocidentais é bastante enfraquecida em relação às demais culturas. No entanto, como demonstram as já mencionadas pesquisas de Haidt — nas quais muitos indivíduos consideram imoral masturbar-se com a carcaça de uma galinha morta antes de comê-la ou limpar o banheiro com a bandeira nacional (Haidt *et al.*, 1993) —, mesmo no Ocidente esse domínio tem relevância. Quando analisamos outras culturas é ainda mais fácil encontrar expressões desse domínio, como, por exemplo, na série de tabus envolvidos no consumo de alimentos e nas práticas sexuais.

No que diz respeito ao domínio ‘*in-group/out-group*’, durante milhões de anos os nossos ancestrais teriam enfrentado o desafio adaptativo de formar e manter coalizões capazes de se defender dos ataques dos grupos rivais. Possivelmente uma longa história de conflitos entre grupos conduziu ao surgimento de sentimentos voltados ao favorecimento de membros participantes do grupo (*in-group*), os quais podem ser facilmente empregados para facilitar o combate contra aqueles que não fazem parte do grupo (Richerson & Boyd, 2005, p. 224). Nesse sentido, mentes tribalistas teriam sido favorecidas pela evolução (Haidt, 2012, cap. 7).

O caráter tribalista de nossas mentes foi evidenciado por um famoso estudo na área de psicologia social. Muzafar Sherif levou um grupo de garotos a um acampamento de verão no parque Robbers Cave, em Oklahoma. Nesse acampamento, ele os dividiu em dois grupos e promoveu jogos e desafios nos quais os grupos eram incentivados a competir. O comportamento tribal cresceu drasticamente com o passar do tempo. Ambos os lados criaram bandeiras e as penduraram em territórios disputados, rasgaram as bandeiras e depredaram os dormitórios dos adversários e fizeram até mesmo armas. Os participantes apreciavam bastante as práticas que promoviam a coesão do grupo, como a elaboração de símbolos que os distinguiam dos demais (Sherif, 1961).

Em outro estudo, Henri Tajfel disse aos participantes que eles estavam participando de um teste sobre julgamento estético. Eles deveriam observar pinturas de Paul Klee e de Wassily Kandinsky e indicar quais eles preferiam. Depois foram separados em dois grupos, supostamente com base nas suas respostas, mas de fato a distribuição foi aleatória. Em seguida os participantes recebiam uma quantia de dinheiro e deveriam dividi-la da forma como quisessem. Eles acabaram favorecendo aqueles que supostamente compartilhavam as mesmas preferências artísticas (Tajfel, 1981; Tajfel *et al.*, 1971). Richerson e Boyd acreditam que a interpretação evolutiva mais plausível para esse resultado é que as pessoas reagem a indicadores de pertencimento a um grupo em razão de que no passado eles serviam para delimitar importantes unidades sociais (Richerson & Boyd, 2005, p. 222). Há evidências de que marcadores simbólicos referentes à delimitação entre diferentes grupos desempenham um papel motivacional importante. Instintos tribais levam as pessoas a empregar marcadores simbólicos para definir o pertencimento a um grupo. Essa definição pode estabelecer quais sujeitos são dignos de empatia, de quais devemos suspeitar e, eventualmente, quais devem ser mortos (Richerson & Boyd, 2005, p. 221).

Pesquisas a respeito do funcionamento dos neurônios-espelho também indicam o caráter tribalista da mente humana. Em seres de nossa espécie, os neurônios-espelho possuem uma forte conexão com áreas do cérebro relacionadas à emoção. Quando vemos alguém sorrir, são ativados alguns dos neurônios que são ativados quando nós mesmos sorrimos. O fato de a outra pessoa estar sorrindo nos faz sentir alegria e aumenta a nossa propensão a rir (Haidt, 2012, cap. 10).

No entanto, uma experiência conduzida pela neurocientista Tania Singer mostra que os nossos neurônios-espelhos não são ativados indiscriminadamente em relação a qualquer pessoa (Singer *et al.*, 2006). O experimento consistia de uma série de jogos envolvendo o dilema do prisioneiro. Cada participante jogava um desses jogos com duas outras pessoas, uma delas colaborava e a outra agia de forma egoísta. Na etapa seguinte, os cérebros do participante e das outras pessoas que participaram do jogo eram escaneados enquanto eles recebiam choques moderados. Os resultados mostraram que os cérebros dos participantes reagiam da mesma maneira quando eles e quando a pessoa que havia colaborado recebiam choques. Eles sentiam empatia e a dor da outra pessoa. Por outro lado, quando o jogador egoísta recebia choques a empatia demonstrada era bem menor e a atividade neural de alguns participantes indicava que eles estavam sentindo prazer. Portanto, nós não sincronizamos a nossa atividade neural com qualquer um. Nós temos uma tendência muito maior de sentir empatia com aqueles que se conformam à nossa matriz moral.

Experiências com a ocitocina apontam na mesma direção. Acredita-se hoje que esse hormônio induz os pais a cuidarem de suas crias e a protegê-las (Churchland, 2011). Experiências com a administração de ocitocina indicam que ela realmente provoca um aumento da cooperação e da preocupação com o bem-estar dos outros indivíduos. Contudo, esse aumento não se verificou em relação a indivíduos que não faziam parte do grupo ao qual o participante analisado pertencia (De Dreu *et al.*, 2010).

Peter Singer cita uma série de exemplos para mostrar como as nossas práticas para evitar o sofrimento são ineficientes em razão de favorecerem de modo desproporcional os membros dos grupos aos quais pertencemos. Estamos dispostos a gastar quantias elevadas para ajudar pessoas que estão de alguma maneira próximas a nós, ainda que o sofrimento delas não seja tão grande, mas não estamos dispostos a gastar quantias muito menores que poderiam salvar diversas vidas em lugares distantes (Singer, 2010). Não hesitamos, por exemplo, em ajudar uma criança ensanguentada e com a perna quebrada de nossa vizinhança, ainda que isso estrague completamente o estofamento novo de nosso carro, mas, como regra, não estamos dispostos a gastar sequer metade do valor que necessitaríamos para reparar o estofamento a fim de salvar um número maior de crianças que estão morrendo de fome do outro lado do mundo (Hauser, 2007, p. 10-11).

A nossa história evolutiva pode ajudar a explicar essa situação. No nosso passado, as oportunidades que tínhamos de prestar auxílio ocorriam ao nosso redor, *e. g.*, um caçador atingido pelo chifre de um búfalo, um membro da família passando fome, uma grávida com problemas. Não havia oportunidades para ajudar os que estavam longe. A psicologia atrelada ao altruísmo evoluiu para lidar com situações como essas, são elas que despertam empatia e desencadeiam reações emocionais capazes de nos motivar a agir (Hauser, 2007, p. 10-11).

Recentemente, em razão dos estudos que conduziu para identificar as diferenças entre conservadores e liberais nos Estados Unidos, Haidt concluiu que talvez fosse mais apropriado dividir o domínio da justiça e da reciprocidade. A partir disso, ele propôs um sexto domínio: liberdade e opressão (*liberty/opression*). Esse domínio estaria mais vinculado à preservação da igualdade, enquanto o domínio da justiça e reciprocidade estaria relacionado à preservação da proporcionalidade nas interações (Haidt, 2012, cap. 8).

A tentativa de Haidt de explicar o surgimento dos domínios morais a partir de uma perspectiva evolucionista enfrenta muitos desafios. As evidências que temos a respeito dos ambientes nos quais ocorreu a maior parte do processo de evolução de nossa espécie são bastante limitadas. Dessa maneira, é possível propor múltiplas histórias que sejam em alguma medida compatíveis com as evidências empíricas existentes, e a tarefa de avaliar qual dessas

histórias é mais plausível envolve um grau elevado de especulação. Para os propósitos do presente capítulo, basta destacar que, ainda que as histórias propostas por Haidt não tenham um embasamento empírico tão robusto quanto desejaríamos acerca da origem dos domínios morais, há inúmeras pesquisas conduzidas por antropólogos e psicólogos que dão suporte à hipótese de que as normas morais existentes nas mais variadas culturas tendem a se agrupar em torno de certos temas. Ou seja, ainda que possa haver controvérsias significativas a respeito da origem desses domínios, da melhor forma de categorizá-los e dos limites que os separam, é muito difícil negar que as normas morais existentes se agrupam em torno de certos temas.

5.2 Críticas à teoria dos domínios morais

Os críticos da teoria dos domínios morais costumam questionar a universalidade dos supostos domínios morais. Jesse Prinz, *e.g.*, contrariamente à ideia da universalidade da proibição de causar dano (*harm*), menciona vários casos envolvendo torturas, guerras e punições corporais (Prinz, 2008, p. 373). Este tipo de argumento não parece relevante contra o modelo defendido por Haidt, pois uma aversão a condutas que causam danos não implica uma proibição a toda e qualquer conduta que cause danos.

Nós não gostamos de ver outras pessoas em situações de aflição e sofrimento; vê-las em tais situações nos causa desconforto, o que demonstra uma predisposição a nos opormos a condutas lesivas. No entanto, Prinz acredita que este tipo de predisposição não é moral. Ele destaca o valor comunicativo do grito de um membro da mesma espécie. O sofrimento dos outros nos alerta para o perigo, é um sinal de que há problema por perto. Por essa razão, não surpreende que ele nos provoque estresse (Prinz, 2008, p. 374). Esse tipo de desconforto, ainda que possa estimular o comportamento pró-social, não seria propriamente moral.

Segundo Prinz, haveria três questões a serem perguntadas: (1) *Os domínios morais são universais?* (2) *Eles podem ser aprendidos?* (3) *Eles são essencialmente morais?* (Prinz, 2008, p. 381). No que diz respeito à primeira pergunta, ele salienta que Haidt admite a variação de ênfase em diferentes culturas em relação aos domínios. Isso traz a possibilidade de que algum dos domínios não seja concebido como moralmente significativo em todas as culturas (Prinz, 2008, p. 381).

Prinz, no entanto, não apresenta evidências antropológicas significativas contrárias à afirmação da universalidade. Ele menciona apenas o caso dos Gahuku Gama de Papua-Nova Guiné. Para esse povo não seria imoral causar danos a outras pessoas, a menos que elas fizessem parte do grupo. De acordo com uma das interpretações possíveis, eles não moralizariam o sofrimento enquanto tal, mas apenas a hierarquia e a reciprocidade. Os Gahuku Gama acreditam

que eles têm responsabilidades para com as pessoas que dependem deles e das quais eles dependem (Prinz, 2008, p. 381-2).

Ainda que ele apresentasse algum caso específico mais forte contra a ideia dos domínios morais, isso não invalidaria a teoria de Haidt. Esses domínios agem em um nível geral influenciando a conduta e estabelecendo vieses. Nada impede que a cultura atue contrabalanceando algum desses domínios. Nas grandes cidades ocidentais, por exemplo, acaba havendo uma ênfase em apenas dois domínios, o relacionado a danos e sofrimento e o relacionado à reciprocidade.

Prinz analisa também a questão do aprendizado. Segundo ele, nós somos dotados de emoções como empatia, respeito, raiva, nojo, *etc.* Essas emoções seriam suficientes para explicar como os domínios morais emergem ao longo do desenvolvimento. Quando uma sociedade, em razão de sentimentos de nojo ou repulsa, moraliza certos comportamentos, podemos afirmar que seus membros possuem um domínio de pureza em sua psicologia moral. Contudo, esse domínio seria apenas uma extensão aprendida de emoções não morais (Prinz, 2008, p. 382).

Por fim, Prinz entende que os domínios não são morais. Eles brotariam de emoções universais que teriam evoluído por diferentes razões e que não teriam sido selecionadas por favorecerem a ocorrência de julgamentos morais. Cada uma dessas emoções teria alguma aplicação não moral. Nós, por exemplo, sentimos empatia pelos doentes, mas não fazemos julgamentos morais sobre eles (Prinz, 2008, p. 382). Segundo Prinz, emoções negativas adquirem significado moral apenas quando aquele que as possui apresenta emoções de culpa contra si mesmo ou contra terceiros (Prinz, 2008, p. 382-3).

De maneira geral, todas as críticas de Prinz à teoria dos domínios morais procuram demonstrar que as capacidades e características envolvidas nos julgamentos morais não são adaptações específicas à moralidade. Como já afirmei, esse tipo de argumento é relevante para a refutação do nativismo moral apenas quando se tem em mente uma concepção evolutiva de inato. Haidt procura explicar a partir de uma perspectiva evolucionista o surgimento das emoções relacionadas à moral e ele certamente considera que essas emoções surgiram como adaptações. Possivelmente, ele acredite que a capacidade para fazer julgamentos morais também é uma adaptação, mas esse não é o aspecto mais relevante de seu modelo de nativismo moral. A teoria de Haidt indica que características humanas canalizadas, sejam elas adaptações ou efeitos secundários de adaptações, garantem que a moralidade se desenvolva e conserve certas características mesmo em ambientes muito distintos. Portanto, as críticas de Prinz não atingem alguns dos aspectos centrais das teorias elaboradas por Haidt.

Defensores da teoria da gramática moral costumam tecer outro tipo de crítica à teoria dos domínios morais. Eles tendem a reconhecer a existência de vieses e a importância das emoções, mas classificam esses fatores como meras interferências na faculdade moral, como algo que pode comprometer a racionalidade dos julgamentos, não como algo intrínseco à moral. Hauser, *e.g.* reconhece que há um viés para favorecer membros do grupo e que ele pode funcionar como uma restrição às possíveis conformações das diversas moralidades. Ele admite que é difícil imaginar o surgimento de normas morais que privilegiem pessoas estranhas ao grupo em detrimento de membros do grupo. No entanto, ele tenta afastar esse viés, como se fosse um empecilho a uma teoria geral da moralidade (Hauser, 2007, p. 133). Teóricos da gramática moral tendem a considerar alguns dos vieses relacionados aos domínios da moral como perturbações (*e.g.*, Hauser, 2007, p. 133). No entanto, ao que tudo indica, esses vieses são inerentes à moral.

Chomsky, ao elaborar a sua teoria linguística, se preocupou em identificar as regras empregadas por um falante ideal em uma comunidade linguística completamente homogênea, *i.e.*, por alguém que não seja afetado por condições gramaticalmente irrelevantes, tais como limitações de memória, distrações e erros (Chomsky, 1965, p. 3). A partir daí, surge a distinção entre competência, *i.e.*, o conhecimento que o falante-ouvinte tem de sua língua, e performance, *i.e.*, o uso real da língua em situações concretas (Chomsky, 1965, p. 4). De modo similar, Mikhail e Hauser procuram aplicar a distinção entre competência e performance em relação à faculdade moral. Para eles certas limitações psicológicas, informações prejudiciais, erros de performance e outros fatores distorceriam os julgamentos morais. A solução seria isolar esses fatores e identificar aqueles julgamentos morais que refletem a operação ideal da faculdade moral (Mikhail, 2011, p. 103). Os proponentes da GMU imaginam uma pessoa idealizada, capaz de empregar a sua faculdade moral sem interferências de outras circunstâncias estranhas a essa faculdade.

Hauser, Mikhail e Dwyer, embora reconheçam a importância das emoções nos julgamentos morais, procuram alcançar em seus experimentos algo como uma situação ideal (*considered judgments*), na qual certos tipos de emoções são evitadas. Eles, assim como Rawls, buscam uma posição neutra, capaz de fornecer julgamentos morais que demonstrem a utilização da faculdade moral sem qualquer distorção (Rawls, 1971; Mikhail, 2011, p. 110).

No âmbito da linguagem, a separação entre faculdade linguística e fatores como a limitação imposta pela memória se mostrou útil e esclarecedora do ponto de vista teórico (Roedder & Harman, 2016, p. 4-5). No entanto, essa separação é problemática no caso da moral, pois não está claro se aquilo que eles consideram interferência não é na verdade constitutivo da

moralidade. Por que deveríamos considerar que a moralidade se restringe a questões de justiça como eles parecem propor?

Há uma dificuldade em combinar a ideia de que os julgamentos morais obedecem uma estrutura gramatical com a ideia de que as emoções desempenham um papel causal na produção das intuições. Mikhail, Hauser e Dwyer, embora reconheçam a existência de uma interface importante entre cognição e emoção, optam por defender que julgamentos morais causam emoções e não o contrário. Diante de evidências empíricas de que podemos manipular os julgamentos morais por meio da manipulação de emoções com a utilização de hipnose e de alterações ambientais (Wheatley & Haidt, 2005; Schnall *et al.*, 2008), eles procuram preservar as suas teorias por meio da distinção entre competência e performance: as emoções afetariam apenas a performance moral, mas não seriam constitutivas da faculdade moral. Os processos emocionais desencadeados antes dos processos computacionais necessários à produção de julgamentos morais poderiam afetar esses últimos e motivar a prática de certas ações, mas não seriam necessários (Dwyer, Huebner & Hauser; 2010, p. 495). Há, no entanto, como vimos no terceiro capítulo, evidências de que respostas emocionais são necessárias à formação da moralidade. A ausência de empatia em crianças com tendências psicopáticas, *e.g.*, interfere no desenvolvimento moral (Blair *et al.*, 2006, 1997; Dupoux & Jacob, 2007, p. 376). Embora os teóricos da GMU pretendam realizar uma abordagem descritiva da realização dos julgamentos morais, eles incluem em suas descrições certa visão normativa. Eles têm uma visão sobre qual deveria ser a moral vigente — essencialmente uma destinada a evitar injustiças (*unfairness*) e a produção de danos e sofrimento (*harm*) — e descartam as outras dimensões envolvidas nos julgamentos morais como se fossem interferências indevidas.

Outra crítica enfrentada por Haidt diz respeito à classificação proposta por ele para as fundações da moral. Suhler e Churchland apontam para outros candidatos, como a modéstia, *e.g.*, que seriam tão qualificados para assumir a posição de ‘fundações’ da moral quanto aqueles admitidos por Haidt. Os mais diversos grupos humanos possuem normas que restringem as ocasiões em que alguém pode vangloriar-se de suas conquistas e de sua posição social — aqueles que costumam gabar-se de seus feitos acabam sendo desprezados e rejeitados pelos seus pares (Suhler & Churchland, 2011, p. 2107). No entanto, o próprio Haidt admite que pode haver outras fundações da moral, além daquelas que ele propõe. Em trabalhos mais recentes (Haidt, 2012), ele adotou uma nova classificação que divide a moralidade em seis fundações, em vez de cinco, e ele não exclui a possibilidade de reconhecer novos domínios. Assim, esse tipo de crítica poderia servir mais como uma forma de aprimoramento da teoria de Haidt do que como uma forma de refutação.

Há, contudo, outras falhas que parecem ser mais relevantes na teoria dos domínios morais. Uma delas diz respeito à vinculação que Haidt faz entre cada um desses domínios (fundações) com certas emoções. Ele associa, *e.g.*, a emoção de raiva ao domínio da reciprocidade. Mas a emoção de raiva pode ser despertada também pela violação de normas pertencentes a outros domínios: quando uma mãe enxerga alguém causando sofrimento a um de seus filhos (*harm/care*), há uma forte reação de hostilidade motivada também pela emoção de raiva. De forma análoga, as outras emoções, como ressentimento, nojo, repulsa, culpa, podem ser recrutadas em favor da imposição de normas pertencentes a domínios diversos. Portanto, não há uma justificativa plausível para associar, de modo isolado, um pequeno grupo de emoções a cada um dos domínios morais (Suhler & Churchland, 2011, p. 2107).

Por fim, analisaremos, nos tópicos seguintes, as duas críticas que nos parecem mais pertinentes em relação à teoria elaborada por Haidt: o fato de ela estar comprometida com uma visão modular acerca da mente humana e a sua incompletude, em razão de não fornecer uma explicação para as etapas mentais prévias ao desencadeamento das intuições.

5.2.1 Modularidade

O conceito de modularidade ganhou bastante notoriedade na Filosofia e na Psicologia desde que Fodor publicou o livro *The Modularity of Mind* (1983). Neste livro, ele defendeu que muitos sistemas subjacentes à percepção e à linguagem são modulares (*low-level systems, input systems*). Por sua vez, adeptos da psicologia evolucionista (Sperber, 2002; Cosmides e Tooby, 1992; Pinker, 2009) elaboraram a tese de que a mente humana seria massivamente modular; ou seja, os sistemas cognitivos mais complexos (*high-level cognition, central systems*), responsáveis por funções como a solução de problemas, a fixação de crenças e planejamento também seriam modulares (Robbins, 2015).

No entanto, o conceito de modularidade tem significados distintos na teoria de Fodor e nas teorias dos psicólogos evolucionistas. Para Fodor, o encapsulamento das informações é uma característica indispensável dos módulos (Fodor, 2000). Isto é, um módulo, enquanto sistema cognitivo encapsulado, não pode ter acesso a informações armazenadas fora do módulo. A ilusão de Müller-Lyer oferece o exemplo clássico dessa propriedade (*Figura 3*): ainda que estejamos convencidos de que as duas linhas centrais possuem o mesmo tamanho, em razão de as termos medido, continuamos a enxergar uma das linhas maior do que a outra em razão do encapsulamento de nosso sistema de percepção visual.

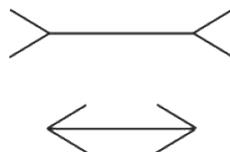


Figura 3: Ilusão de Müller-Lyer (Robbins, 2015)

Já para aqueles que defendem a modularidade massiva, o encapsulamento não é uma característica necessária dos módulos (Robbins, 2015). De fato, eles não poderiam adotar o mesmo conceito de modularidade defendido por Fodor, pois não há como defender de maneira consistente que os sistemas cognitivos centrais sejam encapsulados, pois a fixação de crenças e as atividades de planejamento envolvem a utilização de informações dos mais diversos tipos. Carruthers, defensor do modelo de modularidade massiva, admite a incompatibilidade desse modelo com o conceito de módulo empregado por Fodor:

[If] a thesis of massive mental modularity is to be remotely plausible, then by ‘module’ we cannot mean ‘Fodor-module.’ In particular, the properties of having proprietary transducers, shallow outputs, fast processing, significant innateness or innate channeling, and encapsulation will very likely have to be struck out. That leaves us with the idea that modules might be isolable function-specific processing systems, all or almost all of which are domain specific (in the content sense), whose operations aren’t subject to the will, which are associated with specific neural structures (albeit sometimes spatially dispersed ones), and whose internal operations may be inaccessible to the remainder of cognition (2006, p. 12).

Desta forma, os psicólogos evolucionistas procuram enfatizar a existência de domínios específicos, em vez de encapsulamento. Afirmar que um sistema qualquer é responsável por um domínio específico significa, na linguagem adotada pelos psicólogos evolucionistas, dizer que ele é dedicado exclusivamente ao processamento de uma matéria específica. Nas palavras de Dan Sperber: “A device is domain specific if its function is to process only inputs belonging to some specific empirical domain” (Sperber, 2005, p. 55-6). Nativistas, em geral, costumam invocar mecanismos de domínio específico para explicar a lacuna entre as informações fornecidas pela experiência e as ideias ou crenças que adquirimos a respeito do domínio ao qual pertencem essas informações (Nichols, 2005, p. 354-5). Mecanismos de domínio específico normalmente mencionados incluem aqueles dedicados à linguagem, ao reconhecimento de faces, à leitura de mentes e ao desenvolvimento de um senso comum nas áreas de física e biologia — *folk physics e folk biology* (Hauser, 2007, p. 170 e seguintes; Gelman, 1990; Premack, 1990).

Psicólogos evolucionistas, como regra, consideram que a nossa arquitetura cognitiva é composta por uma série de módulos que teriam evoluído para solucionar os problemas

enfrentados pelos hominíneos ao longo do processo de evolução⁴². A adoção da modularidade é motivada pela observação de que nós resolvemos muitos de nossos problemas diários de maneira eficiente, sem a necessidade de refletirmos ou de empregarmos grandes esforços, e essa eficiência cognitiva demanda algum tipo de explicação. Psicólogos evolucionistas consideram que o ambiente no qual nossos ancestrais viviam apresentava uma série de desafios cognitivos recorrentes, mas que eram quase independentes uns dos outros. Cada um dos agentes de uma tribo precisava, *e.g.*, identificar as ações que violavam as normas de seu grupo. Esse problema é desafiador, pois as regularidades nos comportamentos não são suficientes para resolvê-lo. É impraticável identificar o que é proibido simplesmente observando o que as outras pessoas do grupo não fazem. Dessa maneira, havia grande pressão seletiva no sentido de favorecer os indivíduos que tivessem adaptações específicas capazes de auxiliá-los nessa tarefa. Logo, módulos teriam surgido para auxiliar a resolver cada um dos diferentes tipos de problemas (um módulo para detectar cobras, outro módulo para reconhecer faces, *etc.*). Assim, o aprendizado para solucionar cada um desses desafios seria canalizado e moldado por informações pré-instaladas pertencentes a domínios específicos (Sterelny, 2012).

Sperber emprega uma analogia com a distribuição de funções no corpo humano para defender a tese de que a mente humana é massivamente modular. A necessidade de enfrentar uma série de desafios ao longo da evolução levou os organismos complexos a desenvolverem subsistemas, tais como os diferentes órgãos do corpo humano, responsáveis por executar funções específicas. Como esses subsistemas foram selecionados no decorrer de milhares de anos em razão de serem úteis para solucionar problemas muito distintos um dos outros, e por terem cada um deles uma história filogenética própria, esses subsistemas apresentam muitas diferenças entre si, o que torna difícil a tarefa de elaborar uma definição precisa que englobe todos eles. Segundo Sperber, algo semelhante teria ocorrido no nosso cérebro, ou seja, as estruturas que o compõem foram se diferenciando em módulos ao longo do tempo em razão de favorecerem a solução de problemas específicos enfrentados pelos hominíneos. Como esses módulos apresentam muitas diferenças entre si, em virtude de sua história evolutiva, seria inviável elaborar um conceito rígido de módulo (Sperber, 2005, p. 54).

Haidt, ao elaborar a sua teoria sobre as fundações da moral, adota a tese de que a mente humana é modular, mas rejeita algumas das características propostas por Fodor (1983) para caracterizar a modularidade. Ele adota um modelo inspirado na modularidade massiva proposta por Sperber (2005), que inclui a existência de módulos de aprendizado capazes de produzir

⁴² A descrição do modelo de modularidade feita a seguir é baseada em grande medida na descrição de Kim Sterelny (2012).

outros módulos mais específicos durante o desenvolvimento do indivíduo (Haidt & Joseph 2007, p. 379-80). Ou seja, haveria um módulo de aprendizado anterior a qualquer experiência para cada domínio moral — módulo de primeira ordem. Cada um desses módulos, a partir da experiência adquirida por cada indivíduo, produziria vários outros módulos de trabalho responsáveis por gerar as intuições morais perante situações específicas — módulos de segunda ordem⁴³.

Os módulos de primeira ordem seriam adaptações para uma variedade de fenômenos, ameaças e oportunidades presentes por longos períodos de tempo no ambiente de nossos ancestrais, que teriam como função processar um tipo específico de entrada (*input*). Eles seriam responsáveis por atrair a atenção das pessoas para certos eventos — como demonstrações de crueldade e desrespeito —, e desencadeariam, a partir dos módulos de segunda ordem, reações intuitivas e emoções específicas: empatia, raiva, *etc.* As variações culturais ocorreriam porque as culturas seriam capazes de restringir ou expandir os mecanismos que desencadeiam a atividade de cada um dos módulos de segunda ordem (Haidt, 2012, cap. 6).

Um dos principais motivos que levam Haidt a acreditar na modularidade da mente humana é o fenômeno do *moral dumbfounding*. Em inúmeros casos, os julgamentos morais permanecem inalterados apesar de serem apresentadas novas informações que invalidam as justificativas empregadas para sustentá-los. Haidt acredita que a explicação para esse fato é que os sistemas que produzem esses julgamentos são “em alguma medida encapsulados”, *i.e.*, ao contrário de Fodor, Haidt admite que os módulos tenham acesso a informações armazenadas em outros locais, mas de maneira limitada. Além disso, seguindo a posição normalmente adotadas por psicólogos evolucionistas, Haidt acredita que os módulos dedicados à moralidade pertencem a domínios específicos.

Giroux chama atenção para o fato de que não está muito claro como Haidt chegou à conclusão de que cada um dos cinco domínios da moral corresponderia a um módulo mental diferente (Giroux, 2011, p. 294-5). O simples fato de sermos capazes de classificar as intuições sobre a moral em cinco categorias (*foundations*), não justifica a crença de que cada uma delas seja implementada de forma independente ou por um mecanismo computacional discreto (Mallon, 2008, p. 151). Os domínios propostos por Haidt podem talvez ser universais, mas isso não significa que existam cinco módulos especializados dedicados à moralidade (Giroux, 2011, p. 294-5).

O raciocínio empregado por Haidt para concluir que os supostos módulos dedicados à moralidade são dedicados a domínios específicos segue um modelo de psicologia da ‘caixa-

⁴³ Esta nomenclatura (*second-order modules*) é empregada por Suhler e Churchland (2011, p. 2104)

preta', na medida em que ele leva em consideração apenas relações de entrada (*input*) e saída (*output*). Haidt identifica um padrão de entradas e saídas — estímulos e comportamentos — e conclui que as saídas são o resultado do processamento de um módulo e que esse conjunto de entradas e saídas compõem o domínio ao qual esse módulo é dedicado (Suhler & Churchland, 2011, p. 2106). Um dos problemas desse tipo de raciocínio é que ele não tem grande poder explicativo (ainda que esses módulos existissem, isso não nos diria muita coisa acerca de como eles funcionam) e pode ser utilizado de forma indiscriminada para qualquer processo cognitivo (basta identificar um padrão de entradas e saídas e concluir pela existência de mais um domínio específico). Embora Haidt esteja certo ao afirmar que algumas coisas são mais fáceis de aprender do que outras, esse fato não justifica a inferência de que essa facilidade de aprendizado é fruto da existência de um módulo de aprendizado dedicado a um domínio específico. Poderiam ser elaboradas outras hipóteses que fossem mais consilientes com os resultados de estudos empíricos de outras áreas, como aqueles realizados em neurociência (Suhler & Churchland, 2011, p. 2106).

A ideia de que a mente humana seja composta por diversos mecanismos especializados para a solução de tipos particulares de problemas enfrenta forte oposição. Até mesmo Sperber reconhece que apenas uma minoria dos cientistas cognitivos acredita que a mente seja massivamente modular. Para a maioria deles, a mente é em grande medida não modular. Ainda que admitam a existência de alguns módulos relacionados à percepção, poucos defendem que os sistemas centrais de processamento sejam modulares (Sperber, 2005, p. 53).

Sterelny, *e.g.*, acredita que uma mente massivamente modular não seria capaz de lidar com as grandes variações ambientais enfrentadas pelos hominíneos. Ele chama a atenção para a instabilidade física, biológica e social dos ambientes nos quais eles evoluíram. O tamanho dos grupos, o tipo de divisão de trabalho, a extensão das hierarquias sociais, as interações com outros grupos e espécies, nenhum desses fatores teria permanecido estável. Os humanos enfrentaram os mais diversos desafios ecológicos, como aqueles oferecidos por desertos quentes, pela grande variação sazonal nas proximidades do ártico, pelas florestas tropicais, pela savana africana, e essas diferenças ecológicas interagem com a demografia, com a estrutura social e com a necessidade de especialização, produzindo ambientes muito diversificados. Se a mente humana fosse preponderantemente um conjunto de módulos pré-programados seria muito difícil explicar a grande plasticidade fenotípica demonstrada pelos hominíneos ao lidar com essas diferenças. A nossa mente se desenvolve de formas diferentes, adquirindo diferentes habilidades e capacidades, dependendo do ambiente no qual estamos inseridos. As experiências

culturais canalizam as nossas emoções de diferentes formas. Basta ver a diferença com a qual os insultos são percebidos em diferentes povos (Nisbett & Cohen, 1996).

Sterelny reconhece a alta carga cognitiva dos desafios enfrentados no dia a dia dos hominíneos, mas contesta a premissa de que esses desafios eram estáveis e relativamente independentes. Ele destaca como nós erramos algumas vezes, mesmo que tenhamos sido expostos a uma série de exemplos; mas em outras circunstâncias somos capazes de responder com competência a novos problemas que demandam grande atividade cognitiva. Ele questiona como modelos modulares seriam capazes de explicar as respostas competentes aos problemas novos, pois, à medida que um modelo procura explicar as competências com base em informações pré-instaladas, ele perde poder explicativo em relação à competência demonstrada em face de novos desafios (Sterelny, 2012).

Sperber tenta responder a críticas desse tipo defendendo que o seu sistema de modularidade em dois níveis permitiria a flexibilidade necessária para lidar com as variações ambientais. Embora Sperber sustente que a mente é massivamente modular, a maioria desses módulos não seriam anteriores à aprendizagem. Um conjunto limitado de módulos de aprendizado presentes desde o nascimento propiciaria o surgimento de um número muito maior de módulos de trabalho encarregados de desempenhar as competências cognitivas. O surgimento desses diferentes módulos de segunda ordem dependeria das condições ambientais: diferentes condições provocariam o surgimento de módulos diferentes. Assim, a mente humana teria a capacidade de se moldar, ao longo de seu desenvolvimento, aos mais variados ambientes (Sperber, 2005, p. 57-9).

No entanto, não há consiliência entre o modelo de modularidade proposto por Sperber e adotado por Haidt e os resultados empíricos de pesquisas neurocientíficas. Se a mente humana fosse composta por módulos, deveríamos esperar que a organização do cérebro fosse ao menos compatível com a existência desses módulos (Suhler & Churchland, 2011, p. 2109). No entanto, a anatomia do nosso sistema nervoso central faz com que as ideias de encapsulamento informacional e de especificidade de domínios sejam pouco plausíveis.

Embora as conexões locais presentes no córtex cerebral sejam densas, enquanto as conexões neurais que abrangem distâncias maiores são mais esparsas — *'small world' architecture* —, apenas algumas sinapses separam um determinado neurônio de qualquer outro neurônio presente em nosso cérebro. Como afirmam Suhler e Churchland, “everything is easily accessible to everything else in a few synaptic steps” (Suhler & Churchland, 2011, p. 2109). Esse padrão estaria presente até mesmo no córtex visual primário (V1), a área responsável por receber as entradas da retina através do núcleo geniculado lateral (NGL). Mais de 80% dos

contatos sinápticos presentes em V1 não seriam oriundos do NGL, mas de outras regiões do cérebro (Suhler & Churchland, 2011, p. 2109). Em razão disso, o funcionamento de V1 depende de diversos outros fatores. Niell e Stryker, *e.g.*, demonstraram a influência que o sistema de locomoção exerce no córtex visual primário de ratos: as taxas de disparo dos neurônios de V1 dos animais quando eles estavam correndo era mais do que o dobro da taxa apresentada quando eles estavam em repouso (2010, p. 472). McCauley e Henrich, por sua vez, demonstram como as influências culturais alteram a percepção da ilusão de Müller-Lyer (2006). Resultados como esses desafiam a ideia de que as informações visuais estariam isoladas das informações presentes em partes distantes do cérebro.

A ideia de que os módulos seriam responsáveis por domínios específicos também enfrenta desafios semelhantes. Prinz, *e.g.*, procura demonstrar que os exemplos normalmente mencionados de módulos não correspondem a domínios específicos. Ele aponta a dependência que a leitura de mentes tem em relação à memória de trabalho (Prinz, 2006, p. 28). Prinz menciona a pesquisa conduzida por McKinnon e Moscovitch que demonstrou que o desempenho dos sujeitos cuja memória de trabalho era mantida ocupada era prejudicado em tarefas relacionadas à atribuição de crenças (McKinnon & Moscovitch, 2007). Prinz destaca ainda como os estudos com neuroimagem mostram que a leitura de mentes envolve a utilização de diversas regiões do cérebro, sendo que cada uma dessas regiões contribui também para muitas outras capacidades, ou seja, não são específicas. Da mesma forma, os julgamentos morais recrutam diversas áreas do cérebro, inclusive aqueles centros normalmente associados às emoções (Prinz, 2006, p. 29-30). Além dos estudos mencionados por Prinz, há outros que indicam que até regiões normalmente empregadas na percepção podem ser utilizadas para diversas outras funções. V1, por exemplo, pode servir para decodificar uma língua, solucionar problemas ou realizar simulações mentais (Kosslyn, Ganis & Thompson, 2001; Mellet *et al.*, 2000).

Haidt tem consciência das críticas direcionadas à modularidade. Em virtude disso, ele procura demonstrar que a existência de módulos não é um aspecto central de sua teoria das fundações morais. As fundações que propõe poderiam ser explicadas mais genericamente em termos de “*preparedness*.” A versão mais amena da teoria proposta por Haidt seria a seguinte: a mente humana foi moldada pela evolução de tal modo que as crianças aprendem com grande facilidade a se preocupar com a produção de danos a terceiros (*harm*), com a justiça (*fairness*), com a preservação dos membros de seus respectivos grupos (*in-group*), com a preservação da autoridade e com questões de pureza; no entanto, isso não significa que elas tenham qualquer

conhecimento moral anterior à experiência, apenas que estão preparadas para adquirir certas crenças morais e para resistir a outras (Haidt & Bjorklund, 2008, p. 204).

Quando formulada dessa maneira, a teoria de Haidt aproxima-se muito dos modelos epidemiológicos proposto por Nichols e Sripada, que serão vistos no próximo capítulo. De acordo com Nichols, é plausível que os sistemas emocionais dedicados a reagir perante o sofrimento de outros tenham evoluído como uma forma de superar os desafios ambientais presentes durante o Pleistoceno. No entanto, a influência deles sobre as estruturas cognitivas não seria restrita a domínios específicos, ou seja, as emoções que afetam os julgamentos morais não seriam específicas a esse domínio e poderiam afetar a aquisição de conhecimento em outras áreas: as nossas respostas ao sofrimento alheio poderiam, *e.g.*, afetar a forma como pensamos a respeito de desastres naturais que causam grande sofrimento humano (Nichols, 2005, p. 368). Essa influência difusa das emoções sobre o nosso comportamento, na medida em que indica a inexistência de um encapsulamento, desafia a ideia de que nossa mente possui algum módulo ou conjunto de módulos específicos à moralidade.

O que restou da teoria de Haidt, abandonada a ideia da modularidade, diz em suma que (1) temos mais facilidade em aprender algumas coisas do que outras; e (2) as nossas emoções desempenham um importante papel na determinação daquilo que temos mais facilidade de aprender. Essas afirmações não são muito controversas, pois é difícil imaginar como disposições como as nossas emoções básicas— tristeza, raiva, nojo, medo *etc.*— não interfeririam no aprendizado e no desenvolvimento das normas sociais (Giroux, 2011, p. 289). Todavia, esse tipo de afirmação não é suficiente para esclarecer muitas das nossas dúvidas a respeito do nativismo moral: Haidt afirma que a mente humana é preparada para aprender certas coisas, mas não oferece detalhes a respeito do que constitui essa organização que favorece o aprendizado. Ele não explica (no nível da psicologia cognitiva, da psicologia do desenvolvimento, da neurociência, *etc.*) de que maneira os seres humanos são preparados para a aquisição de normas morais (Suhler & Churchland, 2011, p. 2105). Haidt reafirma algo que já está implícito em diversas pesquisas empíricas sobre o comportamento humano, mas não desvenda os detalhes de como os processos cognitivos de nossa mente resultam na produção de juízos morais. Como veremos no tópico a seguir, a teoria dos domínios morais não dá a devida atenção a elementos cognitivos prévios ao desencadeamento das emoções e intuições.

Antes de passarmos para essa próxima etapa, destacaremos apenas que, de forma geral, as críticas apresentadas nesta seção servem também para a teoria da gramática moral, pois essa teoria também está comprometida com a ideia de que a mente humana é modular. No entanto, ainda que eventualmente a mente humana seja em certa medida modular e, em especial, que a

nossa capacidade linguística seja o resultado do funcionamento de algum módulo, existem diferenças entre a moral e a linguagem que tornam pouco plausível a tese de que nossa capacidade moral seja fruto da operação de determinados módulos.

A presença de reflexividade a partir de processos cognitivos conscientes com consequente alteração das intuições morais contraria a noção de ‘módulos morais.’ Uma das evidências a respeito da modularidade da competência gramatical seria a independência dos julgamentos a respeito da gramaticalidade das sentenças em relação às crenças metalinguísticas (Dupoux & Jacob, 2007, p. 374-5). A descoberta do *moral dumbfounding* foi utilizada com o propósito de sustentar que algo semelhante ocorreria nos julgamentos morais. O fato de haver uma grande dissociação entre os julgamentos morais intuitivos e os processos deliberativos foi interpretado como favorecendo a hipótese de existência de um sistema computacional encapsulado subjacente aos julgamentos morais feitos pelas pessoas em geral— *folk-moral judgment* (Dwyer, Huebner & Hauser; 2010, p. 499). Contudo, os processos de justificação moral também pertencem à capacidade moral humana— não há uma separação tão grande quanto ocorre na linguagem. A presença do *moral dumbfounding* não exclui a possibilidade de que pelo menos alguns julgamentos morais dependam das crenças explícitas que uma pessoa possui sobre a moralidade. É possível que alguém supere a intuição inicial que o leva a favorecer um membro do próprio grupo em nome de uma preferência explícita pela imparcialidade. Se os julgamentos morais podem sofrer a influência de crenças explícitas e de processos conscientes de reflexão, não pode haver um claro isolamento entre a capacidade moral e as crenças morais explícitas (Dupoux & Jacob, 2007, p. 374-5).

5.2.2 Representação mental da ação e teoria da mente

Mikhail e Hauser criticam Haidt por não incluir em seu modelo social-intuicionista uma etapa cognitiva prévia ao desencadeamento de emoções. Nenhuma criatura poderia simplesmente ter uma emoção sem que algo no cérebro houvesse reconhecido que aquela situação era digna de emoção. Haveria uma análise prévia, ainda que inconsciente, responsável por identificar as causas e consequências da ação — quem fez o que, por que, por quais meios, para atingir quais objetivos — e desencadear uma reação emocional. Todo esse processo ocorreria de maneira muito rápida, como quando os nossos cílios detectam a pressão de um objeto, e o nosso cérebro comanda o fechamento das pálpebras (Hauser, 2007, p. 8). Eles destacam que ações muito parecidas ocasionam reações emocionais completamente diversas a depender de como elas são representadas ou percebidas por quem as assiste. Dependendo, por exemplo, das intenções que atribuímos ao autor, teremos reações diversas.

Essa crítica é em certa medida injusta, pois Haidt deixa claro que para ele a distinção relevante é aquela entre dois tipos de cognição — uma rápida (intuitiva) e outra lenta (raciocínio consciente)—, não entre emoção e cognição, pois as emoções envolvem aspectos cognitivos. De qualquer forma, é necessário esclarecer quais processos ocorrem em nossas mentes desde o momento que percebemos uma ação até o momento em que formamos um julgamento moral sobre ela. Hauser e Mikhail parecem estar certos em afirmar que só é possível que uma ação produza determinada emoção após aquele que a percebe tenha formado algum tipo de representação. Portanto, devemos buscar compreender como essa representação, as emoções e outros processos cognitivos mais lentos interagem. Uma explicação completa sobre como ocorrem os julgamentos morais deve esclarecer quais são as estruturas computacionais responsáveis por analisar o cenário em que a ação a ser julgada ocorre (Dwyer, Huebner & Hauser, 2010, p. 494).

Mikhail critica a passagem que Haidt faz em seu modelo social-intuicionista dos estímulos presentes em determinada situação para as intuições despertadas. O problema levantado por Mikhail é que os julgamentos morais não dependem apenas das propriedades superficiais de uma ação. Eles dependem também de um processo avaliativo prévio que determina a forma como essa ação é representada mentalmente, pois o cérebro gera representações da ação que vão além da informação dada (Mikhail, 2011, p. 38, 116 e 309-10).

Mikhail reconhece que as emoções estão claramente ligadas às intuições morais. No entanto, ele considera que o ponto relevante é fazer uma adequada caracterização do sistema de avaliação (*appraisal system*) que essas intuições pressupõem (Mikhail, 2011, p. 122). Apenas emoções não poderiam explicar por que uma ação é considerada errada em um contexto, mas correta em outro. Para que uma emoção surja algo deve provocá-la, algum sistema deve reconhecer uma ação, separá-la em partes e avaliá-la em termos de causas e consequências (Hauser, 2007, p. 52-3). Segundo Hauser é necessário compreender o processo avaliativo que desencadeia as emoções. Esse processo seria a primeira etapa da análise moral e poderia ser a etapa em que os julgamentos morais são formados, ou seja, a parte de nossa mente responsável por avaliar as intenções, ações e consequências poderia ser o centro no qual ocorrem as deliberações morais, a parte de nossa psicologia que julga inicialmente se um comportamento é permissível, obrigatório ou proibido. De acordo com essa visão, as emoções seriam desencadeadas por esses julgamentos morais feitos de maneira inconsciente. Elas não produziriam os julgamentos morais, apenas nos motivariam a agir de determinada forma (Hauser, 2007, p. 30-1).

Segundo Hauser, Haidt explica o fenômeno do *moral dumbfounding* apelando para aquilo que Hauser chama de “criatura humeana.” Esse fenômeno ocorreria por que quando resolvemos dilemas morais não estaríamos raciocinando, mas manifestando intuições decorrentes de emoções. A criatura humeana falaria primeiro e a chamada criatura “kantiana” viria a seguir, simplesmente para oferecer uma racionalização à decisão previamente tomada. Hauser defende que antes que as nossas emoções ou algum tipo de raciocínio consciente possam desempenhar algum papel nos julgamentos morais, deve haver uma etapa prévia de análise estrutural e avaliação das causas e consequências da ação — modelo da dita criatura “rawlsiana” (Hauser, 2007, p. 156). A situação é ainda mais complicada porque há um problema em delimitar claramente a representação e o julgamento moral (*structural dependence on normative Assessments*)⁴⁴.

Os julgamentos morais dependem de uma interpretação da ação por parte daquele que a julga. Quando fazemos esses julgamentos empregamos uma série de inferências sobre as intenções e os estados mentais do autor da ação. A capacidade para fazer essas inferências não pertence apenas ao campo da moral. Nós a empregamos quando interpretamos qualquer ação, por mais estranha que ela possa ser ao campo da moral. Por exemplo, quando vemos alguém cortando um limão e interpretamos que ele está fazendo uma limonada. A interpretação que fazemos de uma determinada ação pode provocar um sentimento de aversão ou de atração. Esse sentimento motiva a adoção de determinado comportamento/reação e explica em boa parte como chegamos a um julgamento moral.

No entanto, a simples reação (*e.g.*, aversão) a determinada circunstância não depende da existência de um julgamento moral. Até mesmo uma ameba é capaz de identificar certas características do ambiente onde se encontra — concentração de íons — e reagir aproximando-se ou afastando-se. Obviamente para que a ameba reaja dessa maneira não é necessária a aplicação de um conjunto de princípios e parâmetros cuja complexidade seja semelhante à de uma gramática. Quando nos deparamos com determinado ambiente ou com determinada ação também reagimos, afastando-nos ou aproximando-nos, conforme as características que identificamos. Contudo, a simples reação emocional a determinada conduta não explica completamente o julgamento moral, pois esse envolve um elemento cognitivo diverso, uma crença a respeito do valor da ação interpretada. Embora seja preciso explicar como essa crença se forma, não há porque imaginar que ela se origine em conformidade com um conjunto de princípios e parâmetros oriundos de uma gramática moral universal.

⁴⁴ Ver Roedder & Harman (2010, p. 279-80).

Portanto, de um lado temos a teoria da gramática moral, que se propõe a explicar os detalhes dos julgamentos morais, mas, em razão das grandes diferenças entre os sistemas envolvidos na produção da linguagem e dos julgamentos morais, falha em oferecer uma explicação plausível. De outro lado, temos a teoria dos domínios morais, que, em sua versão amena, oferece uma explicação plausível para os julgamentos morais, mas falha ao não oferecer uma explicação detalhada. Todavia, talvez a explicação desses julgamentos não envolva muitos detalhes específicos à moralidade; *i.e.*, talvez essa capacidade envolva basicamente uma capacidade geral para reconhecer padrões, o que possibilitaria ao indivíduo identificar as normas de seu grupo, conjugada com uma tendência a moralizar determinados tipos de comportamento. Haveria uma tendência de certos tipos de normas se estabelecerem nas sociedades em razão de predisposições emocionais tipicamente humanas e uma vez estabelecidas essas normas, os novos indivíduos seriam capazes de aprendê-las por mecanismos gerais de aprendizado.

6 Modelos epidemiológicos: Sripada e Nichols

Tendo em vista que as teorias analisadas anteriormente não conseguiram demonstrar a existência de estruturas dedicadas especificamente à moralidade, parece plausível tentar compreendê-la como um fenômeno cultural que se modifica segundo uma dinâmica semelhante em muitos aspectos àquela de outros itens culturais. A seguir analisarei alguns modelos que, para explicar a evolução e o desenvolvimento da moralidade, se inspiram em algumas das teorias naturalistas mais consagradas acerca da evolução cultural: as de Richerson e Boyd e de Sperber. A principal vantagem desse tipo de abordagem em relação ao tipo de explicação usualmente proposto nas ciências sociais é que ele abre espaço para integramos características psicológicas que possuem uma longa história filogenética na explicação da evolução da moralidade. Os modelos que veremos deixam claro que ainda que a capacidade para elaborar juízos morais não seja uma adaptação, isso não significa que podemos explicar o surgimento e a evolução das normas morais sem levarmos em conta a história evolutiva da espécie humana. Esses modelos são menos ‘nativistas’ ou ‘inatistas’ do que os analisados anteriormente, mas ainda podem ser vistos como formas moderadas de nativismo. De acordo com eles, características psicológicas bastante enraizadas em nossa história evolutiva fazem com que a capacidade para julgar moralmente surja em diversos ambientes e canalizam as normas morais, fazendo com que elas tendam a convergir em certas direções.

6.1 Epidemiologia das representações

De acordo com Sperber, em qualquer população humana estão presentes diversas representações, que se dividem em dois tipos. De um lado, temos as *representações mentais*, tais como memórias, crenças, intenções, *etc.*, que são produzidas e utilizadas por um indivíduo e são internas a ele. De outro lado, temos as *representações públicas*, tais como este texto, gestos, vibrações sonoras, artefatos, *etc.*, que são empregadas normalmente como um meio de comunicação entre quem as produz e quem as utiliza, e existem no ambiente que os envolve (Sperber, 1996, p. 32-33).

A maior parte das milhões de representações mentais de um membro de um grupo humano encontram-se apenas em seu cérebro, mas algumas acabam sendo comunicadas a outros integrantes. As representações mentais que um indivíduo constrói ou recupera de sua memória podem fazer com que ele modifique o ambiente físico no qual está inserido, produzindo representações públicas. Essas modificações no ambiente podem levar outros indivíduos a construir novas representações mentais, as quais podem ser armazenadas e, futuramente,

recuperadas. Por sua vez, essas novas representações mentais podem fazer com que esses indivíduos produzam novas modificações ambientais e assim por diante, dando origem a uma corrente causal (Sperber, 1985, p. 77).

Dito de outro modo, representações mentais condicionam quais são os comportamentos adotados por um indivíduo. Alguns desses comportamentos — ou alguns dos resquícios por eles deixados no ambiente — são observados por outros indivíduos. A observação desses comportamentos, ou de seus resquícios, por sua vez, ocasiona o surgimento de novas representações mentais. Algumas vezes, as representações mentais causadas por um comportamento se parecem com aquelas que o causaram. Por meio desse processo, uma representação mental nascida no cérebro de um indivíduo pode ter ‘descendentes’ semelhantes a ela nos cérebros de outros indivíduos. Desse modo, elas podem ser transmitidas de um indivíduo a outro e, assim, se propagar. Sperber chama de *representações culturais* essas representações que, em alguma medida, conseguem se propagar em uma população. Eventualmente, essa propagação pode ser bastante efetiva, de forma que algumas representações mentais, em suas diferentes versões, invadam populações inteiras (1996, p. 1 e p. 25-26).

Para Sperber, a questão fundamental seria explicar por que algumas representações são mais bem-sucedidas do que outras em uma dada população humana; e a resposta a essa questão dependeria de uma compreensão a respeito de como as representações estão distribuídas e de sua dinâmica. Por essa razão, ele propõe que a maneira adequada de compreendermos a evolução cultural é por meio de uma *epidemiologia das representações*. Ele emprega essa expressão para se referir ao estudo das correntes causais que ligam as representações públicas e mentais (a menor dessas cadeias causais seria constituída por um ato de comunicação). O objetivo desse estudo seria explicar os macrofenômenos culturais a partir do efeito cumulativo de dois tipos de micromecanismos: a) os processos intrassubjetivos envolvidos na formação e transformação de representações mentais, tais como pensamento e memória; b) e os processos intersubjetivos de transmissão, pelos quais as representações de um sujeito afetam aquelas de outros sujeitos por meio de modificações no ambiente físico. O primeiro tipo de mecanismo envolveria processos puramente psicológicos, enquanto o segundo envolveria processos parcialmente psicológicos e parcialmente ecológicos, na medida em que envolve tanto o cérebro quanto o ambiente. Em suma, a explicação sobre as causas que fazem com que certas representações se propaguem melhor do que outras em uma dada população humana envolveria a compreensão de duas coisas: como elas são cognitivamente processadas (*cognized*) pelos indivíduos e como elas são comunicadas dentro de um grupo (Sperber, 1985, p. 77; 1996, p. 49-50 e 97-9).

6.1.1 Transmissão das representações: conservação e (re-) produção

Embora Sperber faça analogias entre a propagação de representações mentais e a propagação de agentes infecciosos em uma população, ele enfatiza que, ao contrário do que ocorre na multiplicação de vírus e bactérias, o processo de transmissão de representações, como regra, não ocorre por meio de replicação. Assim, ele rejeita um dos principais pressupostos da memética.

Enquanto vírus e bactérias sofrem mutações de modo ocasional, representações sofreriam, na maior parte das vezes em que são transmitidas, transformações resultantes de processos cognitivos de construção e interpretação (Sperber, 1996, p. 25-26 e 100-1). Segundo ele, a lição mais óbvia das pesquisas em ciências cognitivas é que a recordação de algo não é simplesmente o reverso do processo de armazenamento e que tampouco a compreensão é simplesmente o reverso do processo de expressão, pois “a memória e a comunicação transformam informações” (Sperber, 1996, p. 31).

Sperber utiliza o seguinte exemplo: a compreensão que você tem sobre o que está lendo agora não é, normalmente, uma reprodução na sua mente de meus pensamentos, mas pensamentos construídos por você mesmo que podem ser, mais ou menos, parecidos com os meus. O processo de comunicação seria um processo de transformação que variaria entre dois extremos: a duplicação e a completa perda de informação. Nesse sentido, a replicação de uma representação seria apenas um caso limite, mas de forma alguma a regra (Sperber, 1985, p.75; Sperber, 1996, p.82-83).

Sperber salienta que qualquer um que pretenda descrever de maneira séria representações culturais, tais como textos, canções, flechas, esculturas, *etc.*, como sendo replicações de representações anteriores, deveria estar disposto a indicar de quais representações anteriores elas são réplicas. Entretanto, na maior parte dos casos, elas não são uma simples reprodução de uma única representação ou mesmo de um número limitado delas, mas o resultado de um número indefinido de representações, ainda que algumas delas tenham exercido um papel ‘parental’ mais relevante do que outras (Sperber, 1996, p. 104).

Quando um sujeito adota uma nova representação cultural, ele normalmente não copia simplesmente a variante observada; ele se vale de seus conhecimentos prévios, da sua capacidade de inferência, dos seus interesses e constrói uma nova variante. Essa nova variante provavelmente será diferente da variante observada, tanto porque algumas informações acabam se perdendo no processo, quanto porque o objetivo do agente não é simplesmente adquirir uma réplica, mas alguma habilidade ou conhecimento que sirva às suas preferências (Claidière &

Sperber, 2007, p. 91–92). Sperber menciona como exemplo as crenças que as pessoas têm a respeito do presidente de seu país. A opinião de você a respeito de Temer provavelmente é similar às opiniões de muitas outras pessoas e foi influenciada por algumas delas. Mas é pouco provável que a sua opinião seja uma cópia ou uma média dessas outras posições. É de se esperar que você tenha interpretado as informações sobre Temer a partir de seu conhecimento prévio, de suas preferências, de suas reações afetivas e de sua capacidade de inferência, chegando, assim, à sua atual opinião. Nesse sentido, em casos como esse, a similaridade de sua opinião com a de outras pessoas não seria explicável por um processo de cópia, e sim pela convergência de seus processos cognitivos e afetivos e daqueles de outras pessoas em direção a opiniões psicologicamente atrativas (Sperber, 1996, p. 106).

Apesar de negar que processos de replicação sejam frequentes na evolução cultural, Sperber não nega que processos de seleção darwiniana possam ocorrer com respeito a variantes culturais. Ele acredita na existência de diversos mecanismos e padrões de propagação cultural, assim como ocorre com as doenças, ao explorarem as disposições e suscetibilidades da população hospedeira relevante (Claidière *et al.*, 2014, p. 3).

Desse modo, algumas representações culturais podem, de fato, propagar-se por imitação e outros modos de cópia. Uma criança que quer amarrar os seus sapatos, *e.g.*, tentará imitar fielmente o modelo fornecido por um adulto. De forma semelhante, os monges da Idade Média tentavam reproduzir fielmente os livros que copiavam. Já outras representações se propagam com menos fidelidade: um cozinheiro experiente ao adotar uma nova receita provavelmente a modificará conforme o seu gosto pessoal, em vez de tentar imitá-la estritamente. Portanto, a propagação cultural seria parcialmente preservativa, mas também, em diferentes graus, parcialmente (re-)construtiva. Qual desses processos é mais relevante em determinado caso seria uma questão empírica a ser investigada (Claidière *et al.*, 2014, p. 3-4; Sperber, 1996, p. 102-3).

No que diz respeito aos casos de propagação nos quais há maior fidelidade poderia ocorrer seleção, já nos casos de baixa fidelidade isso não seria possível, pois para que ocorra seleção darwiniana é preciso que a taxa de ‘mutação’ de uma geração para outra fique dentro de certos limites, como salientava George Williams:

The essence of the genetical theory of natural selection is a statistical bias in the relative rates of survival of alternatives (genes, individuals, etc.). The effectiveness of such bias in producing adaptation is contingent on the maintenance of certain quantitative relationships among the operative factors. One necessary condition is that the selected entity must have a high degree of permanence and a low rate of endogenous change, relative to the degree of bias (Williams, 1996, p. 22-3).

Portanto, para que ocorra seleção é preciso que as características que possibilitam a um ente (seja ele um indivíduo biológico ou um item cultural) produzir mais ‘descendentes’ do que

outros entes sejam preservadas: é necessário que as adaptações sejam preservadas de uma geração para outra. Se a taxa de ‘mutação’ for muito grande, isso se torna impossível (Sperber, 1996, p. 102-3).

6.1.2 Atração cultural e fatores de atração

Até aqui vimos que Sperber nega que a transmissão de informação entre os seres humanos, no nível micro, ocorra geralmente por processos de cópia e afirma que a transmissão normalmente implica modificação da informação transmitida. Por outro lado, ele reconhece que, no nível macro, as informações culturais são relativamente estáveis dentro das populações, por vezes, durante gerações — pense em certos ritos funerários e contos infantis, como Chapeuzinho Vermelho, *e.g.*, que já perduram por séculos. A questão que se coloca, então, é como se pode conciliar essas duas posições, uma vez que, em razão de os processos de transmissão não serem fiéis, eles seriam insuficientes para explicar essa macroestabilidade (Claidière & Sperber, 2007, p. 91–2).

Sperber observa que ao formarem representações mentais e públicas os membros de uma população são atraídos para as mesmas direções (Sperber, 1996, p. 118). As variações individuais não se afastam de forma aleatória das representações que lhes servem de modelos; elas tendem a gravitar ao redor de certas posições no espaço de possibilidades, ou seja, as transformações tendem a ser enviesadas na direção de *atratores culturais*. Atratores seriam objetos abstratos similares a centros de gravidade. Afirmar a existência um atrator equivaleria apenas a dizer que as transformações seguem certo padrão e tendem a se agrupar ao redor de um ponto específico. Identificar um atrator não nos dá qualquer explicação causal, apenas sugere o que precisa ser explicado — a distribuição de itens culturais e a sua evolução — e o tipo de explicação causal a ser buscada — a identificação dos *fatores de atração* que causam os vieses nas transformações (Sperber, 1996, p. 108-112; Claidière & Sperber, 2007, p. 91–92).

Atratores decorreriam da existência de diferentes tipos de fatores de atração concretos que afetam a distribuição das variantes culturais. De maneira geral, alguns desses fatores seriam ecológicos ou ambientais (o ambiente inclui aqui tudo o que for externo ao organismo do indivíduo, inclusive todos os outros organismos com os quais interage) e outros seriam psicológicos (Claidière & Sperber, 2007, p. 91–2; Sperber, 1996, p. 84).

Dentre os fatores ecológicos, ele menciona as propriedades demográficas da população, a recorrência de situações nas quais a representação favorece a adoção de uma conduta apropriada, a disponibilidade de mecanismos externos de armazenamento de memória (como a

escrita), a existência de instituições engajadas na transmissão da representação, os tipos de recursos naturais disponíveis, *etc.*

Nesse sentido, *e.g.*, a disponibilidade de substâncias alucinógenas e a prática de consumi-las em uma população são fatores ecológicos que favorecem a propagação do xamanismo, uma vez que o consumo dessas substâncias pode fornecer ‘evidências’ sobre a existência de poderes xamanísticos, reforçando essa instituição. De forma similar, o tipo de peixe presente no ambiente influencia qual tipo de anzol é utilizado pelos pescadores. Nesse caso, o método da tentativa e erro desempenha, sem dúvida, algum papel na escolha do anzol; mas também são relevantes processos construtivos mentais que empregam a imaginação para antecipar e avaliar quais efeitos seriam produzidos pela alteração do tamanho e da forma do anzol. Esse tipo de *variação guiada*⁴⁵ seria apenas um dos processos construtivos relevantes para assegurar a convergência em direção aos atratores (Claidière *et al.*, 2014, p. 4-7; Sperber, 1996, p. 53 e 84).

Sperber salienta que o ambiente humano é, em grande parte, feito pelo próprio homem com base na cultura em que está inserido. Assim, ele reconhece o papel da construção de nichos para a evolução cultural. Para ele, o atual estágio de uma cultura, ou seja, a atual distribuição de representações influencia sobremaneira a futura distribuição de representações: a variação dentro de uma sociedade costuma dirigir-se a conteúdos que são relevantes no contexto das outras representações já presentes (Sperber, 1996, p. 53 e 84).

Por outro lado, dentre os possíveis fatores psicológicos, Sperber menciona o grau de facilidade de aprendizado e de memorização de uma representação particular, a presença de conhecimentos prévios relacionados à representação, a existência de alguma motivação para comunicar o conteúdo da representação, *etc.* (Sperber, 1996, p. 53 e 84).

Em grande parte da literatura das ciências sociais, assume-se que os processos intra- e intersubjetivos possibilitam a fácil circulação de qualquer tipo de representação concebível. A hipótese de que as capacidades cognitivas e comunicativas humanas facilitem ou dificultem a propagação de certas representações costuma ser simplesmente ignorada. No entanto, Sperber chama atenção para o fato de que somos capazes de perceber, por meio de nossa própria experiência pessoal, como a assimilação de representações apresenta diferentes graus de dificuldade: algumas representações, como o teorema da incompletude de Gödel são difíceis de compreender; algumas outras, como um número de vinte dígitos, embora sejam fáceis de compreender, são difíceis de memorizar. Por outro lado, podemos facilmente lembrar uma estória, como a de Chapeuzinho Vermelho, e por diversas vezes ficamos repetindo mentalmente

⁴⁵ Essa expressão foi consagrada por Boyd e Richerson (1988).

um trecho de uma música, mesmo quando queremos muito esquecê-la (pense em músicas como ‘Ai se eu te pego’ ou ‘Yellow submarine’). Sperber então nos pergunta: o que faz com que certas representações sejam mais difíceis de internalizar, recordar e externalizar do que outras? Ele imediatamente rejeita que a complexidade das representações seja a explicação, pois um número de vinte dígitos não é mais complexo do que a estória de Chapeuzinho Vermelho; qualquer computador processa muito mais facilmente o número do que a estória. A complexidade não seria uma explicação, mas algo a ser explicado. O que faz com que algo seja complexo para os seres humanos — *i.e.*, difícil de internalizar, recordar e externalizar — é a forma de organização de nossas capacidades cognitivas e comunicativas. Essas capacidades funcionariam como um filtro que determina quais representações têm maiores chances de difundirem-se em uma dada população humana (Sperber, 1985, p. 79-83).

Em resumo, o ambiente fornece os estímulos (‘inputs’) para os sistemas cognitivos dos membros de uma população e, assim, condiciona a formação e estabilidade das diferentes representações culturais. Já a organização mental dos indivíduos estabelece quais desses estímulos são processados e como eles orientam os comportamentos, os quais, por sua vez, acabam por modificar o ambiente. Dessa maneira, os fatores psicológicos e ecológicos interagem em diferentes níveis, correspondentes às diferentes escalas de tempo: a da evolução biológica, a da história sociocultural, a do desenvolvimento afetivo e cognitivo dos indivíduos e a dos microprocessos de transmissão. Para Sperber, a composição biológica dos seres humanos não apenas torna possível a acumulação cultural, mas também influencia o conteúdo das representações culturais; ou seja, a visão de mundo que temos não é determinada apenas pela história sociocultural. Ele enfatiza que a visão, ainda presente entre muitos cientistas sociais, da mente humana como uma *tabula rasa*, na qual as diferentes culturas inscrevem livremente qualquer conteúdo, é incompatível com o atual estágio de conhecimento nas disciplinas de biologia e psicologia (Sperber, 1996, p. 113).

A abordagem de Sperber apresenta uma grande vantagem em relação àquelas normalmente empregadas nas ciências sociais: ao não atribuir poderes causais a objetos abstratos, *i.e.*, sem existência material, ele cria uma ponte entre a psicologia e as ciências sociais. Tanto os autores ‘materialistas’ quanto os ‘idealistas’ que estudam fenômenos sociais costumam falar de representações tomadas em abstrato (a Religião, a Economia, o Direito) como se elas pudessem entrar em relação causal com o mundo material. Ao agirem assim, eles se comprometem uma ontologia idealista implausível (Sperber, 1985, p. 78).

Sperber, por sua vez, compreende fenômenos culturais como padrões ecológicos de fenômenos psicológicos. Eles não pertenceriam a um nível autônomo de realidade, como os

antirreducionistas defendem, tampouco pertencem apenas à psicologia, como defendem os reducionistas (Sperber, 1985, p. 76). Quando Sperber procura elucidar como um modelo epidemiológico poderia explicar, *e.g.*, a propagação de um mito, ele adota explicitamente uma ontologia materialista. Para ele, o estudo sobre um mito envolveria três tipos de objetos: a) narrativas, *i.e.*, representações públicas que podem ser observadas e gravadas; b) estórias, *i.e.*, representações mentais, que podem ser construídas a partir de narrativas ou expressas por meio delas; c) correntes causais (estórias → narrativas → estórias → narrativas → ...). Cada um desses elementos é um objeto material: cada narrativa é um evento acústico específico; cada estória é um estado cerebral específico; e, logo, a corrente causal que as une também é material. A explicação sobre a existência desse tipo de corrente causal unindo narrativas e estórias exigiria a elaboração de um modelo que levasse em consideração tanto fatores psicológicos — *e.g.*, a forma de organização da memória humana, que faz com que narrativas sejam memorizáveis mais facilmente do que números — quanto ecológicos — *e.g.*, a ausência de escrita em uma determinada população. A interação de fatores como esses poderia ajudar a explicar por que “uma dada narrativa com uma estrutura facilmente memorizável é transmitida com pouca variação em uma tradição oral” (Sperber, 1996, p. 27-29). Assim, a proposta de Sperber fornece um meio para integrarmos os fatores psicológicos nas explicações sobre fenômenos sociais.

A presença dos fatores de atração, psicológicos e ecológicos, nos processos construtivos compensaria as perdas na transmissão. Conforme Sperber, a tarefa de garantir a estabilidade dos conteúdos não precisa ficar a cargo apenas de processos preservativos. Processos construtivos podem manifestar uma tendência de transformar diferentes entradas (‘inputs’) de formas semelhantes, não aleatórias, provocando a convergência das saídas (‘outputs’) em direção aos atratores, assegurando, assim, a estabilidade no nível populacional (Claidière *et al.*, 2014, p. 4; Sperber & Claidière, 2006, p. 289).

6.1.3 Pensamento populacional: Semelhanças entre Sperber e Richerson e Boyd

A abordagem cognitiva de Sperber assemelha-se em muitos pontos à adotada por Richerson e Boyd (2005, 1988). Embora eles provenham de diferentes programas de pesquisa, as suas abordagens vêm convergindo nos últimos anos: o programa cognitivista de Sperber tem prestado cada vez mais atenção à evolução da mente e da cultura; e o programa evolucionista de Richerson e Boyd tem examinado de forma cada vez mais detalhada as bases cognitivas da evolução cultural. Por muito tempo essa convergência não foi reconhecida pelos respectivos proponentes dessas abordagens; em parte, porque os que seguiam algum desses programas

simplesmente ignoravam os autores do outro programa; em parte, em razão de mal-entendidos. Aqui, o próprio Sperber reconhece a sua parcela de culpa, como quando estendeu as críticas dirigidas à memética ao trabalho de Richerson e Boyd, sem atentar para as diferenças relevantes entre essas abordagens (Claidière & Sperber, 2007, p. 89–90; Sperber & Claidière, 2006, p. 283).

Tanto os trabalhos de Sperber quanto os de Richerson e Boyd têm como núcleo o ‘pensamento populacional’ (Richerson & Boyd, 2005, p. 5; Sperber, 1996, p. 118), descrito por Mayr (2000, p. 487) como uma das revoluções fundamentais provocadas por Darwin. O pensamento populacional tem como base a ideia de que fenômenos de grande escala, em um nível macro, podem ser explicados a partir dos efeitos agregados de eventos de pequena escala que ocorrem nas vidas dos organismos. Ele envolve enxergar um sistema (tal como a cultura) como uma população de itens de diferentes tipos relativamente autônomos, cuja frequência se modifica ao longo do tempo. Esses tipos de itens não são definidos em razão de alguma ‘essência’, mas apenas como subpopulações históricas, cujas características podem mudar ao longo do tempo (Lewens, 2015; Claidière *et al.*, 2014, p. 2).

Os modelos de Sperber e os de Richerson e Boyd adotam esse tipo de pensamento, pois enxergam os macrofenômenos culturais como uma população de artefatos ou de itens mentais distribuídos em uma população biológica (normalmente alguma população humana) e em seu habitat, e procuram, então, explicar a dinâmica da distribuição desses itens culturais (Lewens, 2015; Sperber, 1996, p. 2 e p. 118). Eles investigam por que as frequências dos diferentes itens se alteram e por que, em certas circunstâncias, elas se mantêm (Lewens, 2015). Ao realizarem esse tipo de investigação, eles compartilham a meta de explicar a evolução cultural a partir de uma perspectiva que não considera a cultura uma simples extensão da evolução biológica, tal como o faz a sociobiologia, nem como um mero processo estritamente análogo à evolução biológica, tal como o faz a memética (Sperber & Claidière, 2006, p. 283).

O próprio Sperber classifica como epidemiológica a abordagem adotada por Richerson e Boyd para explicar a evolução cultural e salienta que os modelos de coevolução gene-cultura por eles elaborados constituem um dos principais avanços recentes nessa área de pesquisa. Segundo Sperber, o sucesso desses modelos demonstra a importância de se estudar a cultura sob uma perspectiva epidemiológica (Sperber, 1996, p. 3; Claidière *et al.*, 2014, p. 2). Ele reconhece que itens culturais têm estado presentes entre os seres humanos por tempo suficiente para que possam ter influenciado a evolução biológica. No entanto, adverte que a coevolução gene-cultura é um processo lento; logo, incapaz de explicar as mudanças culturais no período histórico mais recente. Os humanos nascem, geração após geração,

com um potencial mental bastante semelhante, mas acabam se desenvolvendo de formas muito distintas em função dos diferentes ambientes físicos e culturais nos quais estão inseridos (Sperber, 1996, p. 114-115). Nesse sentido, ainda que uma abordagem seletcionista, como a adotada em certos modelos de coevolução gene-cultura, possa ser capaz de explicar alguns aspectos culturais, ela é incapaz de explicar mudanças ocorridas em intervalos curtos⁴⁶.

Claidière e Sperber enxergam os processos de seleção como um tipo específico de atração. Eles propõem um diagrama, adaptado da obra de Godfrey-Smith (2009), que distingue três quadros explicativos mais específicos aninhados no quadro populacional de inspiração darwiniana (*Figura 4*). Dentro desse grande quadro, uma população é considerada evolutiva caso as frequências dos diferentes tipos de itens em uma dada etapa sejam explicadas, em grande medida, como uma função dessas frequências em etapas anteriores. Se, além disso, a reprodução dos membros dessa população exibir *variação, herdabilidade e diferenças de aptidão*, então ela estará sujeita à seleção darwiniana. Por fim, se na população em questão, além de estarem preenchidos os requisitos já mencionados, a herdabilidade for assegurada por alguma forma de replicação, então essa dinâmica terá como base a existência de replicadores. Muitas vezes o rótulo ‘darwiniano’ é utilizado apenas em relação a um quadro seletcionista, mas na verdade os quatro quadros são darwinianos em um sentido amplo (Claidière *et al.*, 2014, p. 2; Sperber, 1996, p. 118).

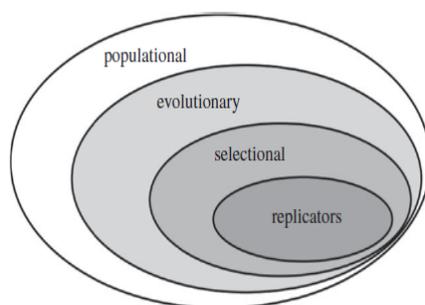


Figura 4 – Quatro quadros explicativos imbricados (Claidière *et al.*, 2014, p. 2)

A descrição dos processos de evolução cultural proposta por Sperber é populacional e evolutiva e pode ser também, dentro de certas circunstâncias, considerada seletcionista. Na visão de Sperber, assim como a seleção com base em replicadores é apenas uma das possíveis formas de seleção, esta é apenas uma das diferentes formas de atração. Assim, ele defende que o

⁴⁶ Os modelos de Sperber e os de Richerson e Boyd apresentam naturalmente algumas diferenças relevantes, as quais são discutidas em detalhe por Sperber e Claidière (2006, p. 284-8). Uma delas é a ênfase que Sperber dá a processos construtivos. Enquanto que para Richerson e Boyd, aparentemente, a *variação guiada* constitui o único processo no qual são introduzidos novos conteúdos de forma não aleatória, Sperber considera que praticamente todos os mecanismos psicológicos envolvidos no aprendizado social combinam processos preservativos e construtivos. Assim, quando alguma representação é transmitida ela também é, normalmente, transformada em alguma medida.

sucesso reprodutivo não é uma alternativa à atração, mas um tipo especial dela: um traço selecionado é um atrator que deve a sua alta frequência ao seu sucesso reprodutivo. Desse modo, não deveríamos perguntar algo como ‘qual é a importância da seleção em relação à atração?’, mas ‘qual parte da atração se deve à seleção?’ A resposta a essa questão dependeria de qual traço e de qual situação específica estamos falando (Claidière *et al.*, 2014, p. 2-7).

6.2 Epidemiologia da moralidade

Shaun Nichols e Chandra Sekhar Sripada, inspirados por Sperber, defendem modelos parecidos com o de Haidt. Para Sripada, a variação dos padrões de moralidade seria mais bem explicada pelo que ele chama de agrupamento temático (*thematic clustering*). Haveria certos temas em torno dos quais a moralidade seria constituída em praticamente todos os grupos humanos, tais como dano, ajuda, compartilhamento, incesto, justiça social e defesa do grupo. Contudo, as regras específicas referentes a esses temas apresentariam bastante variabilidade. A existência desse agrupamento temático — ou, dessa atração, como diria Sperber — explica-se, conforme Sripada, pela existência de vieses atuando sobre os conteúdos das normas morais. A ideia central desse modelo de vieses é que existem certas estruturas psicológicas resultantes do processo de evolução de nossa espécie, como preferências e aversões, que favorecem o surgimento e a manutenção de algumas normas morais. Contudo, esses vieses operam em um nível bastante geral; eles favorecem certos conteúdos temáticos bastante amplos, mas não requerem ou impedem a presença de nenhuma norma moral particular (Sripada, 2008, p. 330-337). Esses vieses podem funcionar influenciando a transmissão de normas morais de um indivíduo para outro, os chamados *Sperberian biases*. Sripada salienta a importância que esses vieses têm, ao longo do tempo, na conformação das normas sociais: “[W]hen their effects are summated over populations and over time, they generate a fairly strong population-level force which can have the effect of changing the distribution of norms in the direction favored by the Sperberian bias” (Sripada, 2008, p. 333).

Alguns tipos de normas podem ser mais ou menos atrativas em razão da forma como elas interagem com nossas preferências, aversões, emoções e outros elementos psicológicos. Por essa razão, algumas normas são mais fáceis de detectar, de inferir, de armazenar e de recordar. Um exemplo de como esse tipo de viés funciona é a forma pela qual o sentimento de nojo e repulsa influencia a transmissão cultural de normas de etiqueta (Sripada, 2008, p. 332-333).

Esses vieses podem também influenciar o surgimento de normas morais (*origination biases*), ou seja, eles podem aumentar a probabilidade de que certos tipos de normas sejam inventadas. Sripada, com base no trabalho de Edward Westermarck, emprega como exemplo a

origem e manutenção de normas proibindo o incesto. Westermarck (2009) estudou a relação entre pessoas que eram criadas juntas em alguns Kibbutzim, mas que não possuíam parentesco entre si. Ele observou que nessas colônias, apesar de não haver normas impedindo que essas pessoas criadas juntas se casassem ou mantivessem relações sexuais, esse tipo de união quase nunca ocorria. Segundo Westermarck, os seres humanos têm um mecanismo que gera aversão a manter relações sexuais com quem eles mantiveram contato íntimo e prolongado durante a infância. A existência dessa aversão funcionaria como um viés, favorecendo o surgimento de normas proibitivas do incesto⁴⁷ (Sripada, 2008, p. 334-6).

Por sua vez, Nichols desenvolveu um modelo chamado de ‘*affective resonance*’, bastante parecido como o modelo dos vieses de Sripada. O modelo de Nichols enfatiza como as disposições emocionais podem atuar restringindo as possibilidades de desenvolvimento da moral. A hipótese formulada por Nichols prevê que, mantidas inalteradas todas as outras circunstâncias, normas que proíbem ações que têm alta probabilidade de despertar emoções negativas possuem maiores chances de ‘sobreviver’ do que normas que não estão conectadas às emoções (Nichols, 2008, p. 270; Nichols, 2004).

Na visão de Nichols, normas não são emoções; tampouco emoções são conceitos envolvidos na semântica presente nos julgamentos morais. As nossas disposições emocionais influenciam os nossos julgamentos morais, mas não devem com eles ser confundidas⁴⁸. No entanto, ele procura explicar como ocorre a coordenação entre emoções e normas. Para isso, ele apela para o papel da evolução cultural. A hipótese dele é que as emoções desempenharam um papel importante na determinação de quais normas sobreviveram ao longo da história. Aquelas normas que proibiam ações propensas a produzir emoções negativas tinham uma aptidão maior⁴⁹. Ele reconhece que há inúmeros fatores que influenciam a evolução cultural, sendo a ressonância afetiva (*affective resonance*) das normas apenas um desses fatores. Os aspectos culturais que são mais salientes do ponto de vista emocional teriam uma capacidade maior de atrair a nossa atenção e de serem memorizados. No caso das normas, aquelas que são suportadas por reações emocionais seriam mais robustas às mudanças temporais (Nichols, 2008, p. 269-70).

As emoções atuariam restringindo quais tipos de normas têm chances significativas de se estabelecerem. Quando uma norma é introduzida, disposições típicas da nossa espécie desempenham um papel importante em definir se ela vai ter sucesso, *i.e.*, ser adotada pela

⁴⁷ É plausível que em certos casos processos culturais contrabalancem vieses inatos. Exemplo disso, seria o declínio de fertilidade em países industrializados (Sripada, 2008, p. 341-2).

⁴⁸ Por essa razão, experiências com crianças que demonstram aversão a praticas danosas (*harm*) devem ser interpretadas com cuidado. O fato de que um bebê manifesta repulsa a determinados comportamentos, não significa que ele esteja condenando moralmente esse comportamento.

⁴⁹ Trata-se da ‘*affective resonance hypothesis*’ (Nichols, 2008, p. 269-70).

sociedade e perdurar no tempo. Portanto, um processo evolutivo estaria em andamento nesse nível e permitiria que apenas as candidatas mais aptas passassem no teste do tempo, e a aptidão delas dependeria da compatibilidade com a condição psicológica da humanidade (Giroux, 2013, p. 181).

De acordo com a hipótese formulada por Shaun Nichols, normas destinadas a evitar a produção de danos e lesões (*harm norms*) teriam uma vantagem em relação às outras no processo de evolução cultural, pois os seres humanos teriam uma forte reação aversiva ao sofrimento de outros. Até mesmo crianças recém-nascidas respondem com aversão a sinais de sofrimento (Simner, 1971). Assim como outras emoções básicas como tristeza, raiva, nojo e medo, reações emocionais ao sofrimento dos outros seriam típicas de nossa espécie (Nichols, 2008, p. 271). As normas destinadas a evitar danos poderiam ter surgido por diferentes razões em diferentes épocas, mas o ponto relevante é que quando elas surgem elas encontram um importante aliado nas emoções. Desse modo, Nichols propõe a seguinte explicação para a aparente universalidade das *harm norms*:

*(i) Harm norms prohibit actions to which we are predisposed to be emotionally averse.
(ii) Norms that prohibit actions to which we are predisposed to be emotionally averse enjoyed enhanced cultural fitness over other norms. If these two claims are right, we should expect harm norms to become widely prevalent, and we thus would have an explanation for the ubiquity of harm norms (Nichols, 2005, p. 363).*

Portanto, assim como Sripada, Nichols procura explicar a presença de certos tipos de normas morais com base na existência de vieses. Contudo, ele ressalta que esses vieses fazem parte de um sistema afetivo, não de informações codificadas em proposições ou de um conjunto de princípios morais (Nichols, 2005, p. 369). Para Nichols, os domínios da moral são gerados a partir de sistemas emocionais, principalmente o sistema afetivo que responde ao sofrimento de outras pessoas. Se esse tipo de modelo estiver correto, as emoções podem desempenhar um relevante papel na conformação das normas morais, o que reduziria o apelo de teorias que propõem a existência de conhecimentos proposicionais não aprendidos (Nichols, 2005, p. 368).

Embora Nichols enfatize aquilo que Haidt chamaria de *harm/care foundation*, ele reconhece a existência de outros candidatos à universalidade no plano moral, como as emoções associadas à justiça (*fairness*) e, em menor medida, a emoção de nojo (Nichols, 2005, p. 356; Nichols, 2008, p. 266). A evolução das normas de etiqueta, por exemplo, comprova como as normas associadas à emoção de nojo possuem uma chance maior de permanecerem inalteradas. Há muita variação em relação às condutas que provocam nojo, mas certas coisas como fluídos corporais, algumas comidas e animais associados a comida estragada desencadeiam o que Haidt chama de ‘*core disgust*’ (Haidt, McCauley & Rozin, 1994).

Norbert Elias descreveu a história das normas de etiqueta no Ocidente, a partir da análise de manuais de etiqueta, desde a Idade Média até o Século XIX (Elias, 1939/2000). Ele menciona vários exemplos de normas que surgiram para proibir ações envolvendo fluídos corporais — *e.g.*, cuspir ou assoar o nariz em certas circunstâncias —, e em cada um desses casos a norma foi preservada. Nichols, por sua vez, fez uma avaliação com base no manual de etiqueta elaborado por Erasmo de Roterdã e verificou que as normas envolvendo ações que provocam nojo tinham uma chance muito maior de sobreviver do que as outras (Nichols, 2002b). Em ambos os casos, os resultados confirmam a hipótese da ressonância afetiva (*affective resonance*), na medida em que as normas associadas a uma resposta emocional demonstraram ter maiores chances de sucesso.

Além disso, várias outras experiências dão suporte ao papel atribuído por Nichols às emoções. No quarto capítulo, já mencionamos o estudo no qual Hamlin, Wynn, e Bloom (2007) demonstraram que crianças com idades entre seis e dez meses apresentam certa aversão ao fantoche que prejudica outro fantoche.

Em outro experimento, Cushman e seus colegas estudaram reações aversivas a condutas violentas. Eles monitoraram os sinais vitais dos participantes enquanto estes simulavam uma série de ações violentas, tais como esmagar a perna de alguém com um martelo. Embora os participantes soubessem que essas ações não causariam qualquer dano, o simples ato de fingi-las causava-lhes uma forte constrição de seus vasos sanguíneos. Os resultados desse estudo salientam a dissociação existente entre o conhecimento explícito sobre as consequências das ações e as respostas afetivas automáticas a elas, tendo em vista que simples propriedades superficiais de ações normalmente danosas foram capazes de desencadear respostas aversivas automáticas (Cushman *et al.*, 2012).

O tipo de valoração intuitiva demonstrada nesses experimentos, atrelada a emoções e feita sem a utilização de raciocínio consciente, ajuda a explicar como as emoções podem funcionar como fatores de atração. Determinadas condutas provocam reações emocionais de aversão, previamente a qualquer julgamento moral. Essas reações emocionais automáticas, que resultam de distintos mecanismos cerebrais e que têm raízes biológicas profundas, servem, então, como base para as normas morais, as quais também sofrem a influência de processos de raciocínio.

Nichols adota uma visão selecionista da evolução cultural e, dessa maneira, se aproxima do pensamento de Richerson e Boyd (1988, 2005). De acordo com esses autores, a estrutura da percepção e da cognição humana faz com que algumas variantes culturais, em virtude de seu conteúdo, sejam aprendidas e lembradas mais facilmente do que outras. Isso faz com que,

ceteris paribus, elas se espalhem mais facilmente. Quando as pessoas adotam preferencialmente certas variantes culturais em vez de outras, temos um processo de transmissão enviesada—*biased transmission* (Richerson & Boyd, 2005, p. 68-70). Os vieses de transmissão surgem em razão da psicologia das pessoas fazer com que elas adotem algumas crenças mais facilmente do que outras. Nesse processo de evolução cultural atuam forças análogas às forças da evolução com base genética, como a seleção, mutação e deriva. Esse tipo de descrição darwiniana da cultura não implica a aceitação dos pressupostos da memética. Não precisamos considerar que a cultura seja divisível em pequenas unidades semelhantes aos genes e que são replicadas fielmente. Trata-se de uma analogia mais frouxa com os genes (Richerson & Boyd, 2005, p. 60).

Além desse tipo de transmissão enviesada, é também possível que processos de seleção no nível do grupo cultural tenham desempenhado papel relevante na distribuição das normas morais: algumas normas podem estar mais presentes, em razão do fato de os grupos que as adotam terem uma aptidão maior. Hipóteses desse tipo podem ter um significativo valor heurístico para a construção de modelos semelhantes aos produzidos por Richerson e Boyd, mas ainda demandam comprovação empírica⁵⁰.

Como já mencionei ao analisar o modelo epidemiológico de Sperber, é possível que determinadas representações culturais, inclusive certas normas morais, possam ser explicadas por modelos selecionistas. Entretanto, isso não significa que esse tipo de explicação sirva para todas as normas morais. Nesse sentido, seria interessante investigar quais são os fatores de atração envolvidos na evolução de diferentes tipos de normas morais. Não há por que imaginar que todas as normas que usualmente classificamos como morais se propaguem seguindo um único padrão.

De qualquer modo, o aspecto central dos modelos desenvolvidos por Haidt, Nichols e Sripada consiste na relevância atribuída a disposições selecionadas ao longo da evolução de nossa espécie, em especial a certos tipos de emoções, na conformação da moralidade existente em um dado grupo humano. Não haveria propriamente algo como um conhecimento moral não aprendido, como no caso da teoria da gramática moral, mas uma tendência a moralizar certos tipos de comportamentos em razão das emoções por eles despertadas e de outras estruturas psicológicas.

É algo que demanda explicação o fato de as normas não se distribuírem de forma aleatória nos diversos grupos humanos. Por que certas normas surgem de forma quase universal, como a proibição de incesto, por exemplo? A ideia de que certas estruturas psicológicas

⁵⁰ Abrantes (2013) discute o possível papel que a seleção em múltiplos níveis, incluindo o nível do grupo, pode ter desempenhado na evolução cultural humana. Em especial, ele examina se uma população de grupos humanos com fenótipos culturais distintos pode ser tratada como uma população darwiniana paradigmática.

decorrentes do processo de evolução humana, como aquelas responsáveis pelas emoções, funcionem como fatores de atração é uma hipótese plausível para explicar isso.

Sripada menciona, além das emoções, também outras estruturas cognitivas associadas às tarefas de detectar, inferir, armazenar e recordar as normas morais, mas, assim como Nichols, se restringe às influências de fatores psicológicos. Apesar disso, a teoria de Sperber também deixa bastante espaço para fatores ambientais em um sentido amplo, incluindo fatores sociais e institucionais, nas explicações acerca da origem e transmissão das normas morais. Uma das grandes contribuições dos teóricos da *construção de nichos* foi demonstrar que os grupos humanos não transmitem apenas os seus genes para as gerações futuras, mas também uma série de modificações no ambiente. Assim, dentre os fatores de atração ambientais, podem existir outros que, assim como as emoções, também estejam enraizados no processo de evolução de nossa espécie. Portanto, uma explicação completa deveria também considerá-los.

Por fim, cabe destacar que possíveis fatores de atração, como as emoções, não precisam ser completamente fixos. O homem compartilha muitos instintos pró-sociais com outros animais que estão próximos a ele na escala evolutiva. Esses instintos envolvem certas emoções compartilhadas com nossos ancestrais e com animais filogeneticamente próximos a nós. Contudo essas emoções não são inalteráveis; elas também sofrem influência em razão do tipo de aprendizado que temos ao longo de nosso desenvolvimento. Até mesmo aquelas emoções consideradas mais básicas sofrem variações decorrentes das diferenças no ambiente cultural. A partir da análise do conceito de inato feita no segundo capítulo, deve ter ficado claro que mesmo essas emoções mais básicas não podem ser compreendidas como algo geneticamente determinado, pois o desenvolvimento de qualquer traço depende de um conjunto de fatores. Aqui é pertinente mencionar de novo o caso dos vegetarianos que, em razão de processos conscientes de raciocínio, passam a sentir nojo em relação ao consumo de carne. Ou seja, mudanças culturais tais como mudanças de crenças e de preferências podem afetar diretamente os estados emocionais dos indivíduos ao longo do seu desenvolvimento (Lewens, 2015).

O exemplo do vegetarianismo indica apenas que a cultura pode modificar o tipo de conduta que dispara o funcionamento do ‘programa’ responsável pela emoção de nojo. No entanto, há estudos que indicam que ela pode alterar também o próprio funcionamento desse programa, *i.e.*, as próprias emoções podem estar sujeitas a variações culturais significativas (Lewens, 2015).

Aqueles que enxergam as emoções básicas como características universais e imunes às diferenças culturais costumam mencionar estudos de Ekman, nos quais ele defende que em todas as culturas as pessoas expressam emoções básicas (alegria, medo, nojo, raiva, surpresa,

tristeza) da mesma maneira (Ekman *et al.*, 1969). No entanto, em um dos primeiros estudos de Ekman, entre os Fore, grupo que habita uma região de Papua Nova Guiné, 56% dos entrevistados classificaram como manifestações de raiva determinadas expressões faciais que ocidentais classificam como manifestações de tristeza; e 45% deles classificaram como medo aquilo que ocidentais classificam como surpresa (Ekman *et al.* 1969). Além disso, o próprio Ekman (1992) enxerga essas emoções básicas como ‘famílias de emoções’ e não considera que elas correspondam a estados afetivos idênticos em todas as culturas.

De todo modo, ainda que as emoções básicas não sejam fixas, elas são mais estáveis do que as normas morais e, assim, podem funcionar como fatores de atração, como reconhece Sperber (1996).

Em suma, embora ainda existam poucos estudos sobre os possíveis fatores de atração relacionados à moralidade, a abordagem epidemiológica se mostra bastante promissora. A epidemiologia das representações de Sperber fornece um quadro bastante amplo, do qual podemos nos servir para buscar explicações para a evolução e desenvolvimento da moralidade. Os modelos de coevolução gene-cultura elaborados por Richerson e Boyd se inserem nesse quadro e podem auxiliar nessa investigação. Modelos matemáticos podem ser elaborados a partir das teorias de Richerson e Boyd e de Sperber, e confrontados com os registros históricos sobre a evolução das normas morais, bem como com simulações de processos de transmissão de normas em laboratório. Estudos desse tipo poderiam ajudar a identificar outros fatores de atração e esclarecer em que medida a evolução da moralidade ocorre segundo processos seletivos.

Conclusão

Os principais programas de pesquisa nativistas analisados ao longo desta dissertação não apresentam boas justificativas para sustentar a existência de estruturas cognitivas dedicadas especificamente à produção de julgamentos morais. Apesar disso, é plausível sustentar que a moralidade é inata, quando se tem em mente um conceito ‘desenvolvimental’ de inato.

No que diz respeito à teoria da gramática moral, há poucas evidências de que os nossos julgamentos morais ocorram segundo princípios universais inacessíveis a nossa consciência. As pesquisas existentes até o momento são insuficientes para afirmar que a moralidade funciona de forma semelhante à linguagem, segundo um modelo de princípios e parâmetros.

Não ficou demonstrado, por exemplo, que a solução para os dilemas envolvendo bondes exige a aplicação de princípios universais. A comparação com casos paradigmáticos pode ser suficiente para explicar as escolhas dos entrevistados. A ausência de justificativas plausíveis por parte deles reforça a tese de que a moralidade é fundamentalmente intuitiva, não de que existam princípios universais inacessíveis à consciência.

Até mesmo Hauser, Dwyer e Huebner (Hauser, 2008; Dwyer, Huebner & Hauser; 2010, p. 491, Hauser, 2007, p. 48) admitem que o estado atual da pesquisa no âmbito da moralidade é muito pior do que aquele encontrado por Chomsky e por outros linguistas que começaram a elaborar a teoria da gramática generativa no final dos anos 50. Enquanto há décadas linguistas catalogam os detalhes das línguas mundiais e dissecam os padrões de ordem das palavras e de concordância, não há nada equivalente no âmbito da moral. Em face da ausência de uma descrição rica da competência moral dos adultos, a analogia com a linguagem fica comprometida.

No entanto, a adoção de uma versão fraca dessa analogia pode ser útil para orientar pesquisas futuras. Há, de fato, certas semelhanças entre linguagem e moralidade que podem ser exploradas, usando o caso da linguagem como ferramenta heurística na abordagem dos problemas da moralidade.

A analogia linguística baseia-se na ideia de que há uma lacuna entre os estímulos que recebemos e o desenvolvimento da competência moral. No entanto, ainda que essa lacuna exista, e a nossa mente seja capaz de superá-la em razão de estar de alguma forma preparada para o aprendizado moral, isso não implica a existência de princípios não aprendidos e inacessíveis à nossa consciência.

Contudo, a capacidade para fazer julgamentos morais surge muito cedo, e de forma robusta, em diversos ambientes. Essa facilidade de aprendizado sugere que o cérebro humano está de alguma forma preparado para aprender normas morais, embora não saibamos o quanto.

Muitos dos que se opõem ao nativismo moral concordam que a aquisição de normas é preparada biologicamente. Entretanto, eles acreditam que as adaptações cruciais são perceptuais e motivacionais, e não específicas à moralidade. Ou seja, eles adotam uma concepção evolutiva do caráter inato. Se adotassem uma concepção desenvolvimental, concluiriam que a moralidade, ou pelo menos grande parte dos componentes que a constituem, é inata. Prinz, por exemplo, considera a moralidade um efeito secundário de traços psicológicos desenvolvidos para outros propósitos. Ele reconhece que ela é restringida por nossas características biológicas e que não nascemos como uma *tabula rasa*. Nossas emoções, nossa capacidade para atribuir estados mentais e o cuidado que temos por nossos parentes funcionariam como blocos de construção da moralidade, segundo ele, mas não se confundiriam com ela (Prinz, 2013).

Não é controverso que os nossos traços biológicos influenciam os tipos de normas que adquirimos — quase todos os envolvidos nesse debate concordariam que os elementos emotivos e comportamentais que influenciam a moralidade foram selecionados. O que está em questão é a existência de certas capacidades cognitivas específicas, como pensar sobre questões morais ou emitir julgamentos.

O desenvolvimento da capacidade para fazer julgamentos morais segundo um calendário similar em diferentes sociedades também indica tratar-se de uma capacidade que, para manifestar-se, depende da existência de cultura, mas não é redutível a ela. As normas morais existentes nas diferentes comunidades que conhecemos agrupam-se em torno dos mesmos temas. Esses fatos constituem evidência forte em favor da existência de uma base biológica responsável por produzir certos vieses no aprendizado moral. Portanto, aqueles que utilizam o conceito desenvolvimental de inato estão justificados em afirmar que a moralidade é, em alguma medida, inata. No entanto, os estudos realizados até o momento não permitem identificar com precisão em que medida essa canalização resulta de fatores genéticos ou de fatores ambientais, tais como a construção de nichos.

Para refutar o argumento da universalidade seria necessário provar que a moralidade é algo fácil de ser aprendido, e esse não parece ser o caso. Não obstante, embora a moralidade seja um fenômeno universal que costuma organizar-se em torno de certos temas, as normas morais apresentam grande variação entre os diferentes grupos culturais. Essa é mais uma das dificuldades apresentadas em relação ao modelo de princípios e parâmetros.

Quanto à teoria dos domínios morais, proposta por Haidt, ela tem o mérito de incluir certas características psicológicas típicas da espécie humanas, tais como as nossas disposições emocionais, na explicação sobre a nossa capacidade para julgar moralmente. No entanto, ela também é vulnerável a críticas importantes.

Em primeiro lugar, ao aceitar a tese da modularidade massiva, Haidt compromete-se com uma visão a respeito da mente humana que não é consiliente com os resultados de pesquisas neurocientíficas. Apesar disso, Haidt busca deixar claro que esse não é o aspecto mais relevante do seu modelo de nativismo. Em razão de ter ciência das críticas à modularidade massiva, ele propõe subsidiariamente uma versão mais amena de sua teoria. Conforme essa versão, os domínios morais poderiam ser explicados mais genericamente em termos de preparação (*preparedness*). Haidt sustenta que características humanas canalizadas — sejam elas adaptações ou efeitos secundários de adaptações — garantem que a moralidade se desenvolva e conserve certas características mesmo em ambientes muito distintos.

A teoria de Haidt sofre ainda críticas em razão de sua incompletude. Ela não fornece uma explicação para as etapas mentais prévias ao desencadeamento das intuições e, nesse sentido, parece ignorar que só é possível que uma ação produza determinada intuição após aquele que a percebe tenha formado algum tipo de representação. Portanto, em sua versão amena, a teoria dos domínios morais oferece uma explicação plausível para os julgamentos morais, mas falha ao não oferecer uma explicação a respeito do papel que inferências sobre as intenções e os estados mentais do autor da ação desempenham no desencadeamento das intuições morais.

Com relação aos modelos epidemiológicos, eles retomam a ideia, já presente na teoria de Haidt, de que algumas normas interagem melhor com nossas estruturas cognitivas, preferências, aversões, emoções, *etc.*, e explicam como essa relação entre normas e características psicológicas pode provocar o surgimento de vieses que favorecem a propagação de certas normas em detrimento de outras. A ideia de que características psicológicas decorrentes do processo de evolução humana funcionem como fatores de atração é uma hipótese plausível para explicar a recorrência, em diversas culturas, de normas similares. Desse modo, tendo em vista que fatores de atração parecem canalizar a moralidade, poderíamos dizer que ela é em alguma medida inata, ao menos nesse sentido ‘desenvolvimental.’

Embora sejam plausíveis, as explicações propostas por Sripada e Nichols são ainda insuficientes, pois se restringem às influências de fatores psicológicos. A teoria da epidemiologia das representações formulada por Sperber deixa claro que a atração envolve tantos fatores psicológicos quanto ecológicos. Assim, seria imprescindível identificar quais os

fatores de atração ecológicos mais relevantes nos processos de origem e transmissão das normas morais. Para a execução dessa tarefa, teóricos da construção de nichos, como Sterelny, oferecem contribuições relevantes, tendo em vista que no caso humano o papel do ambiente construído culturalmente parece se sobressair em relação ao ambiente ‘natural.’

Além disso, os modelos epidemiológicos desenvolvidos até agora são também, assim como a teoria de Haidt, suscetíveis à crítica de não darem a devida atenção a elementos cognitivos prévios ao desencadeamento das emoções e intuições.

Apesar dessas lacunas, a abordagem epidemiológica se mostra a mais promissora dentre as analisadas. Por meio dela, a fim de compreendermos o desenvolvimento e a evolução da moralidade, podemos nos valer de todo arcabouço teórico desenvolvido para o estudo da evolução cultural.

Autores como Sripada e Nichols se valem dos resultados obtidos em modelos consagrados sobre a evolução cultural, para afirmar que mesmo vieses relativamente fracos podem ter consequências muito relevantes quando os seus efeitos se somam ao longo de diversas gerações. Desta forma, eles se valem de modelos evolutivos para explicar por que adquirimos determinadas normas, integrando assim a explicação sobre o desenvolvimento da moralidade à explicação sobre a evolução das normas morais.

O quadro da epidemiologia das representações proposto por Sperber e dentro do qual se inserem as propostas de Nichols e Sripada deixa espaço tanto para modelos selecionistas quanto para modelos que sejam evolutivos apenas em um sentido mais amplo. Essa característica da epidemiologia das representações de Sperber é vantajosa, pois não há razão para pressupormos *a priori* que normas morais de diferentes tipos se propaguem da mesma forma. Tampouco há estudos suficientes para sabermos quais formas de propagação são predominantes.

Ao longo deste trabalho, além de enfatizar o papel de nossas disposições emocionais, mencionei também diversas outras capacidades envolvidas nos julgamentos morais: reconhecimento de padrões, leitura de mentes, consciência, linguagem, memória, emoções, etc. Cada uma delas é empregada na formulação de julgamentos morais, mas não é dedicada apenas a essa tarefa. Tendo em vista que dispomos de todas essas capacidades, o fato de julgarmos moralmente deixa de ser algo tão surpreendente. A autoconsciência, como já disse, é capaz de explicar a capacidade de pensar normativamente pelo menos em um sentido prudencial. Obviamente essa explicação não é tão satisfatória como gostaríamos, uma vez que a explicação para a autoconsciência continua sendo uma questão bastante controversa.

Outra questão que ainda precisa ser respondida diz respeito à causa que leva tantas pessoas a tratarem as normas morais como se fossem objetivas. A explicação para isso

provavelmente se relaciona ao fato de que frequentemente tentamos utilizar as normas morais a fim de avançar nossos interesses. Elas são vistas como uma espécie de quadro comum, ao qual podemos recorrer para solucionar disputas. Nesse contexto, o apelo a normas que suspostamente existem independentemente da vontade e dos desejos dos envolvidos pode parecer mais persuasivo. Do ponto de vista cognitivo, o julgamento moral parece seguir o mesmo padrão de outros julgamentos. A diferença é que aquelas normas que qualificamos como morais costumam estar associadas a certas reações emocionais.

Referências

- ABARBANELL, L.; HAUSER, M. D. Mayan morality: an exploration of permissible harms. *Cognition*, v. 115, n. 2, p. 207–224, maio 2010.
- ABRANTES, P. C. C. Human evolution and transitions in individuality. *Contrastes. Suplemento*, n. 18, p. 203–220, 2013.
- ABRANTES, P. C. C. Conflito e cooperação na evolução humana. *Ciência & Ambiente*, v. 48, p. 289–301, 2014a.
- ABRANTES, P. C. C. Naturalizando a epistemologia. In: ABRANTES, P. C. C. (Org.). *Epistemologia e Cognição*. Brasília: Editora Universidade de Brasília, 1994. p. 171–218. Disponível em: <<http://repositorio.unb.br/handle/10482/13492>>. Acesso em: 29 out. 2016.
- ABRANTES, P. C. C. Natureza e cultura. *Ciência & Ambiente*, v. 48, p. 7–21, 2014b.
- ALMEIDA, F. P. L. DE. *A evolução da mente normativa: origens da cooperação humana*. Dissertação (Mestrado em Filosofia). Brasília: Universidade de Brasília, 2011. Disponível em: <<http://repositorio.unb.br/handle/10482/9867>>. Acesso em: 10 nov. 2016.
- ALMEIDA, F. P. L.; ABRANTES, P. C. C. A teoria da dupla herança e a evolução da moralidade. *Principia: an international journal of epistemology*, v. 16, n. 1, ago. 2012. Disponível em: <<http://www.periodicos.ufsc.br/index.php/principia/article/view/23635>>.
- ANDERSON, S. W. *et al.* Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, v. 2, n. 11, p. 1032–1037, nov. 1999.
- ARIEW, A. Innateness. In: MATTHEN, M.; STEPHENS, C. (Org.). *Philosophy of Biology*. 1 ed. Amsterdam; London: North Holland, 2007. p. 567–585.
- ARIEW, A. Innateness and canalization. *Philosophy of Science*, v. 63, p. S19–S27, 1996.
- ARIEW, A. Innateness Is Canalization: In Defense of a Developmental Account of Innateness. In: HARDCASTLE, V. (Org.). *Biology Meets Psychology: Conjectures, Connections, Constraints*. MIT Press, Cambridge, MA, 1999. p. 117–138. Disponível em: <<http://cogprints.org/332/>>. Acesso em: 30 out. 2016.
- ARMSTRONG, S. L.; GLEITMAN, L. R.; GLEITMAN, H. What some concepts might not be. *Cognition*, v. 13, n. 3, p. 263–308, maio 1983.
- ATRAN, S. Folk biology and the anthropology of science: cognitive universals and cultural particulars. *The Behavioral and Brain Sciences*, v. 21, n. 4, p. 547-569-609, ago. 1998.
- AUGSBERG, S. Der „naturalistische Fehlschluß“ als juristische Argumentationsfigur. *ARSP: Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, v. 94, n. 4, p. 461–476, 2008.
- AXELROD, R. *The Evolution of Cooperation*. New York: Basic Books, 1984.
- AYALA, F. J. Colloquium paper: the difference of being human: morality. *Proceedings of the National Academy of Sciences of the United States of America*, v. 107 Suppl, p. 9015–9022, 2010.
- AYER, A. J. *Language, Truth and Logic*. 2. ed. Dover Publications Inc., 2002.
- BAILLARGEON, R. Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology*, v. 23, n. 5, p. 655–664, 1987.
- BAIN, A. *Mental and moral science*. Andesite Press, 2015.
- BAKER, L. R. *Naturalism and the First-Person Perspective*. 1. ed. Oxford; New York: Oxford University Press, 2013.
- BARSALOU, L. The Instability of Graded Structure: Implications for the Nature of Concepts. In: NEISSER, U. (Org.). *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge University Press, 1987. p. 101–140.
- BARTHOLOW, B. D.; DICKTER, C. L.; SESTIR, M. A. Stereotype activation and control of race bias: cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology*, v. 90, n. 2, p. 272–287, fev. 2006.
- BIRNBACHER, D. *Analytische Einführung in die Ethik*. De Gruyter, 2003.

- BISCHOF, N. *Moral*. 1. ed. Wien: Böhlau Köln, 2012.
- BLAIR, R. *et al.* The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, v. 34, n. 2, p. 192–198, 1997.
- BLAIR, R. J. R. The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, v. 11, n. 9, p. 387–392, set. 2007.
- BLAIR, R. J. R. *et al.* The development of psychopathy. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, v. 47, n. 3–4, p. 262–276, abr. 2006.
- BOEHM, C. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, Mass.: Harvard University Press, 1999.
- BOURGUIGNON, E.; UCKO, L. G. *Diversity and Homogeneity in World Societies*. New Haven, Conn.: Human Relations Area Files, 1973.
- BOYD, R.; RICHERSON, P. J. *Culture and the Evolutionary Process*. 2. ed. University Of Chicago Press, 1988.
- BROWN, D. E. *Human universals*. Philadelphia: Temple University Press, 1991.
- BROWN, R.; HANLON, C. Derivational complexity and order of acquisition in child speech. In: HAYES, J. R. (Org.). *Cognition and the development of language*. New York: Wiley, 1970. .
- BRUNI, T.; MAMELI, M.; RINI, R. A. The Science of Morality and its Normative Implications. *Neuroethics*, v. 7, n. 2, p. 159–172, 2014.
- CAREY, S. *The Origin of Concepts*. Reprint ed. Oxford; New York: Oxford University Press, 2011.
- CARNAP, R. *Philosophy and Logical Syntax*. New York: Ams Pr Inc, 1979.
- CARRUTHERS, P. *The Architecture of the Mind*. 1 ed. Oxford : Oxford ; New York: Clarendon Press, 2006.
- CASHDAN, E. Hunters and gatherers: Economic behavior in bands. In: PLATTNER, S. (Org.). *Economic anthropology*. Stanford, Calif: Stanford University Press, 1989. .
- CASTRO, E. V. DE. Os pronomes cosmológicos e o perspectivismo ameríndio. *Mana*, v. 2, n. 2, p. 115–144, out. 1996.
- CHOMSKY, N. *Aspects of the Theory of Syntax*. Cambridge: The MIT Press, 1965.
- CHOMSKY, N. *Knowledge of Language: Its Nature, Origins, and Use*. New York: Praeger, 1986.
- CHOMSKY, N. *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, Mass: The MIT Press, 1987.
- CHOMSKY, N. *Syntactic Structures*. Walter de Gruyter, 2002.
- CHOMSKY, N. *The Logical Structure of Linguistic Theory*. New York: Springer, 1975.
- CHURCHLAND, P. S. *Braintrust: What Neuroscience Tells Us about Morality*. Princeton, N.J.: Princeton University Press, 2011.
- CIARAMELLI, E. *et al.* Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, v. 2, n. 2, p. 84–92, jun. 2007.
- CLAIDIÈRE, N.; SCOTT-PHILLIPS, T. C.; SPERBER, D. How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, v. 369, n. 1642, p. 20130368, 19 maio 2014.
- CLAIDIÈRE, N.; SPERBER, D. The role of attraction in cultural evolution. *Journal of Cognition and Culture*, v. 7, n. 1, p. 89–111, 1 mar. 2007.
- COLLINS, J. Faculty Disputes. *Mind & Language*, v. 19, n. 5, p. 503–533, 1 nov. 2004.
- COSMIDES, L.; TOOBY, J. Cognitive adaptations for social exchange. In: BARKOW, J. H.; COSMIDES, L.; TOOBY, J. (Org.). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, 1992. p. 163–228.
- COURTENAY-SMITH, N. *A truly special love story: Two married thalidomide survivors living happily 50 years after drug's launch*. Disponível em:

- <<http://www.dailymail.co.uk/news/article-561360/A-truly-special-love-story-Two-married-thalidomide-survivors-living-happily-50-years-drugs-launch.html>>. Acesso em: 25 out. 2016.
- CUSHMAN, F. *et al.* Simulating murder: The aversion to harmful action. *Emotion*, v. 12, n. 1, p. 2–7, 2012.
- CUSHMAN, F.; YOUNG, L.; GREENE, J. D. Multi-system moral psychology. In: DORIS, J. M. (Org.). *The handbook of moral psychology*. Oxford; New York: Oxford University Press, 2010. p. 47–71.
- CUSHMAN, F.; YOUNG, L.; HAUSER, M. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science*, v. 17, n. 12, p. 1082–1089, dez. 2006.
- DAMASIO, A. *Descartes' Error: Emotion, Reason, and the Human Brain*. Reprint ed. London: Penguin Books, 2008.
- DAMASIO, A. R. *Self Comes to Mind: Constructing the Conscious Brain*. Pantheon Books, 2010.
- DANCY, J. Moral Particularism. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Fall 2013 ed. Disponível em: <<http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>>. Acesso em: 30 out. 2016.
- DANIELS, N. On Some Methods of Ethics and Linguistics. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, v. 37, n. 1, p. 21–36, 1980.
- DARWIN, C. *The Descent of Man and Selection in Relation to Sex*. 1. ed. New York: Cambridge University Press, 2009. v. 1.
- DE WAAL, F. *Good natured: The origins of right and wrong in humans and other animals*. Cambridge: Harvard University Press, 1996.
- DECASPER, A. J.; SPENCE, M. J. Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, v. 9, n. 2, p. 133–150, 1 abr. 1986.
- DEL-BEN, C. M. Neurobiologia do transtorno de personalidade anti-social. *Revista de Psiquiatria Clínica*, v. 32, n. 1, p. 27–36, 2005.
- DENNETT, D. C. *The Intentional Stance*. Rev. ed. Cambridge, Mass.: A Bradford Book, 1989.
- DESCARTES, R. *Meditações Sobre Filosofia Primeira*. 1. ed. Campinas: Unicamp, 2004.
- DREU, C. K. W. D. *et al.* The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans. *Science*, v. 328, n. 5984, p. 1408–1411, 11 jun. 2010.
- DUPOUX, E.; JACOB, P. Universal moral grammar: a critical appraisal. *Trends in Cognitive Sciences*, v. 11, n. 9, p. 373–378, 2007.
- DWYER, S. How Good Is the Linguistic Analogy? In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind: Culture and Cognition*. New York: Oxford University Press, 2007. v. 2. p. 238–255.
- DWYER, S. How Not to Argue that Morality Isn't Innate: Comments on Prinz. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*. Cambridge: MIT press, 2008. v. 1. p. 407–418.
- DWYER, S. Moral dumbfounding and the linguistic analogy: Methodological implications for the study of moral judgment. *Mind and Language*, v. 24, n. 3, p. 274–296, 2009.
- 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, jul. 2010.
- EDGERTON, R. B. *Sick Societies: Challenging the Myth of Primitive Harmony*. 1. ed. New York : Toronto : New York: Free Press, 1992.
- EISENBERG, N.; LOSOYA, S.; SPINRADD, T. Affect and prosocial responding. In: DAVIDSON, R. J.; SHERER, K. R.; GOLDSMITH, H. H. (Org.). *Handbook of Affective Sciences*. 1. ed. New York: Oxford University Press, 2003. p. 787–803.
- EKMAN, P. An argument for basic emotions. *Cognition & Emotion*, v. 6, n. 3, p. 169–200, 1 maio 1992.

- EKMAN, P.; SORENSON, E. R.; FRIESEN, W. V. Pan-cultural elements in facial displays of emotion. *Science (New York, N.Y.)*, v. 164, n. 3875, p. 86–88, 4 abr. 1969.
- ELIAS, N. *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. Revised ed. Oxford ; Malden, Mass: Blackwell Publishing, 2000.
- ESKINE, K. J.; KACINIK, N. A.; PRINZ, J. J. A bad taste in the mouth: gustatory disgust influences moral judgment. *Psychological Science*, v. 22, n. 3, p. 295–299, mar. 2011.
- ESTES, W. K. Array models for category learning. *Cognitive Psychology*, v. 18, n. 4, p. 500–549, out. 1986.
- FESSLER, D. M. T. *et al.* Disgust sensitivity and meat consumption: A test of an emotivist account of moral vegetarianism. *Appetite*, v. 41, n. 1, p. 31–41, 2003.
- FIELD, T. M.; WALDEN, T. A. Production and discrimination of facial expressions by preschool children. *Child Development*, v. 53, n. 5, p. 1299–1311, out. 1982.
- FISKE, A. P. *Human Sociality*. Disponível em: <<http://www.sscnet.ucla.edu/anthro/faculty/fiske/reloadModel.htm>>. Acesso em: 31 out. 2016.
- FISKE, A. P. The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, v. 99, n. 4, p. 689–723, 1992.
- FITCH, W. T. *The Evolution of Language*. 1. ed. Cambridge ; New York: Cambridge University Press, 2010.
- FITZPATRICK, W. Morality and Evolutionary Biology. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Spring 2016 ed. Disponível em: <<http://plato.stanford.edu/archives/spr2016/entries/morality-biology/>>. Acesso em: 20 jul. 2016.
- FODOR, J. A. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. 1. ed. Cambridge, Mass.: A Bradford Book, 2001.
- FODOR, J. A. *The modularity of mind: an essay on faculty psychology*. Cambridge: MIT Press, 1983.
- FOOT, P. Morality as a System of Hypothetical Imperatives. *The Philosophical Review*, v. 81, n. 3, p. 305–316, 1972.
- FRANKENA, W. K. The Concept of Morality. *Journal of Philosophy*, v. 63, n. 21, p. 688–696, 1966.
- FRANKENA, W. K. The Naturalistic Fallacy. *Mind*, v. 48, n. 192, p. 464–477, 1939.
- GABENNESCH, H. The Perception of Social Conventionality by Children and Adults. *Child Development*, v. 61, n. 6, p. 2047–2059, 1990.
- GASKING, D. Clusters. *Australasian Journal of Philosophy*, v. 38, n. 1, p. 1–36, 1 maio 1960.
- GAZZANIGA, M. S. *The Social Brain: Discovering the Networks of the Mind*. 1. ed. New York: Basic Books, 1985.
- GEACH, P. T. Assertion. *Philosophical Review*, v. 74, n. 4, p. 449–465, 1965.
- GELMAN, R. First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, v. 14, n. 1, p. 79–106, jan. 1990.
- GERT, B.; GERT, J. The Definition of Morality. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Spring 2016 ed. Disponível em: <<http://plato.stanford.edu/archives/spr2016/entries/morality-definition/>>. Acesso em: 30 out. 2016.
- GIBBARD, A. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press, 2003.
- GIBBARD, A. *Wise Choices, Apt Feelings - A Theory of Normative Judgement*. Revised ed. Oxford: OUP Oxford, 2002.
- GIROUX, J. The origin of moral norms: A moderate nativist account. *Dialogue*, v. 50, n. 2, p. 281–306, 2011.
- GIROUX, J. The Origin of Moral Norms and the Role of Innate Dispositions. In: MUSSCHENGA, B.; HARKAMP, A. VAN (Org.). *What Makes Us Moral? On the*

- capacities and conditions for being moral*. Library of Ethics and Applied Philosophy. Springer Netherlands, 2013. p. 169–186. Disponível em: <http://link.springer.com/chapter/10.1007/978-94-007-6343-2_11>. Acesso em: 30 out. 2016.
- GODFREY-SMITH, P. *Darwinian Populations and Natural Selection*. Oxford; New York: Oxford University Press, 2009.
- GODFREY-SMITH, P. Genes and codes: lessons from the philosophy of mind. In: HARDCASTLE, V. G. (Org.). *Where Biology Meets Psychology: Philosophical Essays*. 1. ed. Cambridge, Mass: A Bradford Book, 1999. p. 305–331.
- GREENE, J. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. 1. ed. New York: Penguin Press, 2013.
- GREENE, J. D. *et al.* An fMRI investigation of emotional engagement in moral judgment. *Science*, v. 293, n. 5537, p. 2105–8, 2001.
- GREENE, J. D. Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics*, v. 124, n. 4, p. 695–726, 1 jul. 2014.
- GREENE, J. D. *et al.* Cognitive Load Selectively Interferes with Utilitarian Moral Judgment. *Cognition*, v. 107, n. 3, p. 1144–1154, jun. 2008.
- GREENE, J. D. *et al.* The neural bases of cognitive conflict and control in moral judgment. *Neuron*, v. 44, n. 2, p. 389–400, 14 out. 2004.
- GREENE, J.; HAIDT, J. How (and where) does moral judgment work? *Trends in Cognitive Sciences*, v. 6, n. 12, p. 517–523, 1 dez. 2002.
- GREENWALD, A. G.; MCGHEE, D. E.; SCHWARTZ, J. L. Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, v. 74, n. 6, p. 1464–1480, jun. 1998.
- GRIFFITHS, P. Our Plastic Nature. In: GISSIS, S. B.; JABLONKA, E. (Org.). *Transformations of Lamarckism*. Cambridge, Mass.: The MIT Press, 2011. p. 319–330. Disponível em: <<http://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262015141.001.0001/upso-9780262015141-chapter-30>>. Acesso em: 30 out. 2016.
- GRIFFITHS, P. The Distinction Between Innate and Acquired Characteristics. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Fall 2009 ed. Disponível em: <<http://plato.stanford.edu/archives/fall2009/entries/innate-acquired/>>. Acesso em: 30 out. 2016.
- GRIFFITHS, P. E.; GRAY, R. D. Developmental Systems and Evolutionary Explanation. *The Journal of Philosophy*, v. 91, n. 6, p. 277–304, 1994.
- GRIFFITHS, P.; MACHERY, E.; LINQUIST, S. The Vernacular Concept of Innateness. *Mind & Language*, v. 24, n. 5, p. 605–630, 1 nov. 2009.
- HAIDT, J. Moral Psychology and the Law: How Intuitions Drive Reasoning, Judgment, and the Search for Evidence. *Alabama Law Review*, v. 64, p. 867–903, 2013.
- 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.
- HAIDT, J. *The righteous mind: why good people are divided by politics and religion*. 1. ed. New York: Pantheon Books, 2012.
- HAIDT, J.; BJORKLUND, F. Social Intuitionists Answer Six Questions About Morality. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral Psychology, The Cognitive science of morality: Intuition and diversity*. Cambridge: MIT Press, 2008. v. 2. p. 181–217.
- HAIDT, J.; JOSEPH, C. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, v. 133, n. 4, p. 55–66, 2004.
- HAIDT, J.; JOSEPH, C. The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules. In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind*. New York: Oxford University Press, 2007. v. 3. p. 367–392.

- Haidt, J.; Kesebir, S. Morality. In: Fiske, S.; Gilbert, D.; Lindzey, G. (Org.). *Handbook of Social Psychology*. 5. ed. Hoboken, NJ: Wiley, 2010. p. 797–832.
- Haidt, J.; Koller, S. H.; Dias, M. G. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, v. 65, n. 4, p. 613–628, out. 1993.
- Haidt, J.; McCauley, C.; Rozin, P. Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences*, v. 16, n. 5, p. 701–713, maio 1994.
- Hamlin, J. K.; Wynn, K.; Bloom, P. Social evaluation by preverbal infants. *Nature*, v. 450, n. 7169, p. 557–559, nov. 2007.
- Hare, R. M. *Freedom and Reason*. Oxford University Press, 1977.
- Hare, R. M. *Moral Thinking: Its Levels, Methods and Point*. 1. ed. Oxford; New York: Oxford University Press, 1982.
- Hare, R. M. *The Language of Morals*. Reprint ed. London; New York: Oxford University Press, 1991.
- Harmann, G. *Explaining Value: and Other Essays in Moral Philosophy*. 1. ed. New York: Clarendon Press, 2000.
- Harmann, G. Using a linguistic analogy to study morality. In: Sinnott-Armstrong, W. (Org.). *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*. Cambridge: MIT Press, 2008. v. 1. p. 345–352.
- Harmann, G.; Mason, K.; Sinnott-Armstrong, W. Moral Reasoning. In: Doris, J. M. (Org.). *The Moral Psychology Handbook*. 1. ed. Oxford; New York: Oxford University Press, 2010. p. 206–245.
- Hauser, M. D. *Moral minds: the nature of right and wrong*. New York: Harper Collins, 2007.
- Hauser, M.; Young, L.; Cushman, F. Reviving Rawls’ linguistic analogy: Operative principles and the causal structure of moral actions. In: Sinnott-Armstrong, W. (Org.). *Moral Psychology, The Cognitive Science of Morality: Intuition and Diversity*. Cambridge: MIT Press, 2008. v. 2. p. 107–143.
- Henrich, J. et al. “Economic man” in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *The Behavioral and Brain Sciences*, v. 28, n. 6, p. 795-815-855, dez. 2005.
- Henrich, J. et al. In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *The American Economic Review*, v. 91, n. 2, p. 73–78, 2001.
- Hoerster, N. Zum Problem der Ableitung eines Sollens aus einem Sein in der analytischen Moralphilosophie. *ARSP: Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, v. 55, n. 1, p. 11–39, 1969.
- Hoffman, M. L. *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press, 2001.
- Horgan, T.; Timmons, M. Troubles for New Wave Moral Semantics: The “Open Question Argument” Revived. *Philosophical Papers*, v. 21, n. 3, p. 153–175, 1992.
- Hume, D. *A Treatise of Human Nature*. Waiheke Island: The Floating Press, 2009.
- Hume, D. *An Enquiry Concerning the Principles of Morals*. Jovian Press, 2016.
- Jackendoff, R. *Patterns In The Mind: Language And Human Nature*. New York, NY: Basic Books, 1994.
- Johnston, T. D. The persistence of dichotomies in the study of behavioral development. *Developmental Review*, v. 7, n. 2, p. 149–182, 1987.
- Joyce, R. Evolution and Moral Naturalism. In: Clark, K. J. (Org.). *The Blackwell Companion to Naturalism*. John Wiley & Sons, Inc, 2016. p. 369–385.
- Joyce, R. Moral Anti-Realism. In: Zalta, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Fall 2015 ed. Disponível em: <http://plato.stanford.edu/archives/fall2015/entries/moral-anti-realism/>. Acesso em: 20 jul. 2016.

- JOYCE, R. *The evolution of morality*. Cambridge: MIT Press, 2006.
- JOYCE, R. The many moral nativisms. In: STERELNY, K. *et al.* (Org.). *Cooperation and its Evolution*. Cambridge: MIT Press, 2013. p. 549–572.
- JOYCE, R. The origins of moral judgment. *Behaviour*, v. 151, n. 2–3, p. 261–278, 2014.
- KAHNEMAN, D.; FREDERICK, S. A model of heuristic judgment. In: HOLYOAK, K. J.; MORRISON, R. G. (Org.). *The Cambridge Handbook of Thinking and Reasoning*. New York: Cambridge University Press, 2005. p. 267–293.
- KANT, I. *Grundlegung zur Metaphysik der Sitten*. Hamburg: Meiner, F, 1999.
- KELLY, D. *et al.* Harm, Affect, and the Moral/Conventional Distinction. *Mind & Language*, v. 22, n. 2, p. 117–131, abr. 2007.
- KELLY, D.; MACHERY, E.; MALLON, R. Race and Racial Cognition. In: DORIS, J. M. (Org.). *The Moral Psychology Handbook*. Oxford; New York: Oxford University Press, 2010. p. 433–472.
- KELLY, D.; STICH, S. Two Theories About the Cognitive Architecture Underlying Morality. In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind: Volume 3: Foundations and the Future*. New York: Oxford University Press, 2007. p. 348–366.
- KIEHL, K. A. A cognitive neuroscience perspective on psychopathy: evidence for paralimbic system dysfunction. *Psychiatry Research*, v. 142, n. 2–3, p. 107–128, 15 jun. 2006.
- KILLEN, M.; SMETANA, J. G. (Org.). *Handbook of Moral Development*. 1. ed. Mahwah, N.J.: Psychology Press, 2005.
- KINZLER, K. D.; DUPOUX, E.; SPELKE, E. S. The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, n. 30, p. 12577–12580, 24 jul. 2007.
- KITCHER, P. Battling the undead: How (and how not) to resist genetic determinism. *Thinking about evolution: Historical, philosophical, and political perspectives*, v. 2, p. 396–414, 2001.
- KOENIGS, M. *et al.* Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, v. 446, n. 7138, p. 908–911, 19 abr. 2007.
- KORSGAARD, C. Morality and the distinctiveness of human action. In: MACEDO, S.; OBER, J. (Org.). *Primates and philosophers: How morality evolved*. Princeton: Princeton University Press, 2006. p. 98–119.
- KORSGAARD, C. M. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. 1. ed. Oxford ; New York: Oxford University Press, 2008.
- KOSSLYN, S. M.; GANIS, G.; THOMPSON, W. L. Neural foundations of imagery. *Nature Reviews. Neuroscience*, v. 2, n. 9, p. 635–642, set. 2001.
- KREBS, D. The evolution of morality. In: BUSS, D. M. (Org.). *The Handbook of Evolutionary Psychology*. 1. ed. Hoboken, N.J: Wiley, 2005. .
- KRIPKE, S. A. *Naming and Necessity*. 1. ed. Cambridge, USA: Harvard University Press, 2001.
- LEHRMAN, D. S. A critique of Konrad Lorenz's theory of instinctive behavior. *The Quarterly Review of Biology*, v. 28, n. 4, p. 337–363, dez. 1953.
- LEITER, B. Naturalism in Legal Philosophy. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Fall 2014 ed. Disponível em: <<http://plato.stanford.edu/archives/fall2014/entries/lawphil-naturalism/>>. Acesso em: 29 out. 2016.
- LENMAN, J. Moral Naturalism. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Spring 2014 ed. Disponível em: <<http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/>>. Acesso em: 29 out. 2016.
- LENNOX, J. G. Darwin was a teleologist. *Biology and Philosophy*, v. 8, n. 4, p. 409–421, 1993.
- LERNER, J. S.; GOLDBERG, J. H.; TETLOCK, P. E. Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility. *Personality and Social Psychology Bulletin*, v. 24, n. 6, p. 563–574, 6 jan. 1998.

- LEVINE, R. A.; CAMPBELL, D. T. *Ethnocentrism: Theories of Conflict, Ethnic Attitudes, and Group Behavior*. New York: John Wiley & Sons, 1972.
- LEWENS, T. *Cultural Evolution: Conceptual Challenges*. 1. ed. Oxford, United Kingdom: Oxford University Press, 2015.
- LINQUIST, S. *et al.* Exploring the folkbiological conception of human nature. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, v. 366, n. 1563, p. 444–453, 12 fev. 2011.
- LOCKE, J. *An Essay Concerning Human Understanding*. Amherst, N.Y.: Prometheus Books, 1995.
- LOMBROZO, T. The Role of Moral Commitments in Moral Judgment. *Cognitive Science*, v. 33, n. 2, p. 273–286, 1 mar. 2009.
- LUCO, A. The Definition of Morality: Threading the Needle. *Social Theory and Practice*, v. 40, n. 3, p. 361–387, 2014.
- LUO, Q. *et al.* The neural basis of implicit moral attitude--an IAT study using event-related fMRI. *NeuroImage*, v. 30, n. 4, p. 1449–1457, 1 maio 2006.
- MACHERY, E.; MALLON, R. The evolution of morality. In: DORIS, J. M. (Org.). *The Moral Psychology Handbook*. 1. ed. Oxford ; New York: Oxford University Press, 2010. p. 3–46.
- MACINTYRE, A. What Morality is Not. *Philosophy*, v. 32, n. 123, p. 325–335, 1957.
- MACKIE, J. L. *Ethics: Inventing Right and Wrong*. New ed. Penguin, 1990.
- MALLON, R. Reviving Rawls's linguistic analogy inside and out. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral Psychology, The Cognitive Science of Morality: Intuition and Diversity*. Cambridge: MIT Press, 2008. v. 2. p. 145–156.
- MALLON, R.; NICHOLS, S. Dual Processes and Moral Rules. *Emotion Review*, v. 3, n. 3, p. 284–285, 7 jan. 2011.
- MALLON, R.; NICHOLS, S. Rules. In: DORIS, J. M. (Org.). *The Moral Psychology Handbook*. 1. ed. Oxford ; New York: Oxford University Press, 2010. p. 433–472.
- MALT, B. C.; SMITH, E. E. Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, v. 23, n. 2, p. 250–269, 1 abr. 1984.
- 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, 18 set. 2013.
- MAMELI, M.; BATESON, P. An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 366, n. 1563, p. 436–443, 12 fev. 2011.
- MAMELI, M.; BATESON, P. Innateness and the Sciences. *Biology and Philosophy*, v. 21, n. 2, p. 155–188, 2006.
- MARGOLIS, E.; LAURENCE, S. Concepts. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Spring 2014 ed. Disponível em: <<http://plato.stanford.edu/archives/spr2014/entries/concepts/>>. Acesso em: 26 out. 2016.
- MAYR, E. Teleological and Teleonomic, a New Analysis. In: COHEN, R. S.; WARTOFSKY, M. W. (Org.). *A Portrait of Twenty-five Years*. Boston Studies in the Philosophy of Science. Springer Netherlands, 1974. p. 133–159. Disponível em: <http://link.springer.com/chapter/10.1007/978-94-009-5345-1_10>. Acesso em: 27 out. 2016.
- MAYR, E. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Eleventh printing ed. Cambridge, Mass.: Belknap Press, 2000.
- MCCAULEY, R. N.; HENRICH, J. Susceptibility to the Müller-Lyer Illusion, Theory-Neutral Observation, and the Diachronic Penetrability of the Visual Input System. *Philosophical Psychology*, v. 19, n. 1, p. 79–101, fev. 2006.
- MCDOWELL, J.; MCFETRIDGE, I. G. Are Moral Requirements Hypothetical Imperatives? *Proceedings of the Aristotelian Society, Supplementary Volumes*, v. 52, p. 13–42, 1978.

- MCKINNON, M. C.; MOSCOVITCH, M. Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, v. 102, n. 2, p. 179–218, 2007.
- MEDIN, D. L.; SCHAFFER, M. M. Context theory of classification learning. *Psychological Review*, v. 85, n. 3, p. 207–238, 1978.
- MEDIN, D. L.; SMITH, E. E. Concepts and Concept Formation. *Annual Review of Psychology*, v. 35, n. 1, p. 113–138, 1984.
- MELLET, E. *et al.* Functional anatomy of high-resolution visual mental imagery. *Journal of Cognitive Neuroscience*, v. 12, n. 1, p. 98–109, jan. 2000.
- MIKHAIL, J. M. *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge: Cambridge University Press, 2011.
- MIKHAIL, J. M. The Poverty of the Moral Stimulus. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*. Cambridge: MIT press, 2008. v. 1. p. 367–406.
- MIKHAIL, J. M. Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, v. 11, n. 4, p. 143–152, abr. 2007.
- MILGRAM, S. *Obedience to Authority: An Experimental View*. New York: Harper Perennial Modern Classics, 1974.
- MINTON, C.; KAGAN, J.; LEVINE, J. A. Maternal Control and Obedience in the Two-Year-Old. *Child Development*, v. 42, n. 6, p. 1873–1894, 1971.
- MOLL, J.; DE OLIVEIRA-SOUZA, R.; BRAMATI, I. E.; *et al.* Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, v. 16, n. 3 Pt 1, p. 696–703, 2002.
- MOLL, J.; DE OLIVEIRA-SOUZA, R.; ESLINGER, P. J.; *et al.* The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, v. 22, n. 7, p. 2730–2736, 2002.
- MONAHAN, J. L.; MURPHY, S. T.; ZAJONC, R. B. Subliminal mere exposure: specific, general, and diffuse effects. *Psychological Science*, v. 11, n. 6, p. 462–466, nov. 2000.
- MOORE, G. E. A reply to my critics. In: SCHILPP, P. A. *The philosophy of G. E. Moore*. 3. ed. London: Cambridge University Press, 1968. v. 2. p. 554–611.
- MOORE, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1922.
- NICHOLS, S. Innateness and Moral Psychology. In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind: Structure and Contents*. New York: Oxford University Press, 2005. v. 1. p. 353–370.
- NICHOLS, S. Norms with feeling: towards a psychological account of moral judgment. *Cognition*, v. 84, n. 2, p. 221–236, jun. 2002a.
- NICHOLS, S. On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution. *Philosophy of Science*, v. 69, n. 2, p. 234–255, 2002b.
- NICHOLS, S. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford; New York: Oxford University Press, 2004.
- NICHOLS, S. Sentimentalism naturalized. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral Psychology, The Cognitive Science of Morality: Intuition and Diversity*. Cambridge: MIT Press, 2008. v. 2. p. 255–274.
- NIELL, C. M.; STRYKER, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, v. 65, n. 4, p. 472–479, 25 fev. 2010.
- NIELSEN, C. F. It's Complicated – Moral Nativism, Moral Input and Moral Development. In: MUSSCHENGA, B.; HARSKAMP, A. VAN (Org.). *What Makes Us Moral? On the capacities and conditions for being moral*. Library of Ethics and Applied Philosophy. Springer Netherlands, 2013. p. 187–206. Disponível em: <http://link.springer.com/chapter/10.1007/978-94-007-6343-2_11>. Acesso em: 30 out. 2016.

- NISBETT, R. E.; COHEN, D. *Culture Of Honor: The Psychology Of Violence In The South*. 1. ed. Boulder, Colo: Westview Press, 1996.
- NOSEK, B. A.; GREENWALD, A. G.; BANAJI, M. R. The Implicit Association Test at Age 7: A Methodological and Conceptual Review. *Social psychology and the unconscious: The automaticity of higher mental processes*. Frontiers of social psychology. New York, NY, US: Psychology Press, 2007. p. 265–292.
- NUCCI, L. P.; TURIEL, E.; ENCARNACION-GAWRYCH, G. Children'S Social Interactions and Social Concepts Analyses of Morality and Convention in the Virgin Islands. *Journal of Cross-Cultural Psychology*, v. 14, n. 4, p. 469–487, 12 jan. 1983.
- NUCCI, L.; TURIEL, E. God's Word, Religious Rules, and Their Relation to Christian and Jewish Children's Concepts of Morality. *Child Development*, v. 64, n. 5, p. 1475–1491, 1993.
- PENNINGTON, J.; SCHLENKER, B. R. Accountability for Consequential Decisions: Justifying Ethical Judgments to Audiences. *Personality and Social Psychology Bulletin*, v. 25, n. 9, p. 1067–1081, 1 set. 1999.
- PINKER, S. *How the Mind Works*. New York: W. W. Norton, 2009.
- PIZARRO, D. A.; BLOOM, P. The intelligence of the moral intuitions: comment on Haidt (2001). *Psychological Review*, v. 110, n. 1, p. 193–198, jan. 2003.
- PIZARRO, D. A.; UHLMANN, E.; BLOOM, P. Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, v. 39, n. 6, p. 653–660, 2003.
- PLATO. *Meno and Other Dialogues*. Oxford: Oxford University Press, 2005.
- PREMACK, D. The infant's theory of self-propelled objects. *Cognition*, v. 36, n. 1, p. 1–16, jul. 1990.
- PRINZ, J. *Beyond human nature: how culture and experience shape our lives*. Allen Lane, an imprint of Penguin Books, 2012.
- PRINZ, J. Is morality innate. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*. Cambridge: MIT Press, 2008. v. 1. p. 367–406.
- PRINZ, J. Is the mind really modular. In: STAINTON, R. J. (Org.). *Contemporary debates in cognitive science*. Oxford: Blackwell Publishing, 2006. p. 22–36.
- PRINZ, J. Where Do Morals Come From?—A Plea for a Cultural Approach. In: CHRISTEN, M. et al. (Org.). *Empirically Informed Ethics: morality between facts and norms*. Berlin: Springer, 2013. p. 99–118.
- PRINZ, J.; NICHOLS, S. Moral Emotions. In: DORIS, J. M. (Org.). *The handbook of moral psychology*. Oxford; New York: Oxford University Press, 2010. p. 111–146.
- RAWLS, J. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press, 1971.
- RICHESON, P. J.; BOYD, R. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University Of Chicago Press, 2005.
- RICHESON, J. A. et al. An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, v. 6, n. 12, p. 1323–1328, dez. 2003.
- RICHESON, J. A.; SHELTON, J. N. When Prejudice Does Not Pay Effects of Interracial Contact on Executive Function. *Psychological Science*, v. 14, n. 3, p. 287–290, 1 maio 2003.
- RIDGE, M. Moral Non-Naturalism. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Fall 2014 ed. Disponível em: <<http://plato.stanford.edu/archives/fall2014/entries/moral-non-naturalism/>>. Acesso em: 29 out. 2016.
- RIPS, L. J.; SHOBEN, E. J.; SMITH, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, v. 12, n. 1, p. 1–20, 1 fev. 1973.
- RITCHIE, J. *Understanding Naturalism*. 1. ed. Stocksfield England: Acumen Pub. Ltd., 2008.
- ROBBINS, P. Modularity of Mind. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Summer 2015 ed. Disponível em:

- <<http://plato.stanford.edu/archives/sum2015/entries/modularity-mind/>>. Acesso em: 31 out. 2016.
- ROEDDER, E.; HARMAN, G. Linguistics and moral theory. In: DORIS, J. M. (Org.). *The Moral Psychology Handbook*. Oxford: Oxford University Press, 2010. p. 273–296.
- ROEDDER, E.; HARMAN, G. *Moral Grammar*. Disponível em: <http://www.princeton.edu/~harman/Papers/Moral_Grammar.pdf>. Acesso em: 31 out. 2016.
- ROSALDO, M. Z. *Knowledge and Passion*. Cambridge, Eng.; New York: Cambridge University Press, 1980.
- ROSAS, A. Ética evolucionista: o enfoque adaptacionista da cooperação humana. In: ABRANTES, P. C. (Org.). *Filosofia da Biologia*. Porto Alegre: Artmed Editora, 2011. p. 296–314.
- ROSCH, E. Principles of Categorization. In: MARGOLIS, E.; LAURENCE, S. (Org.). *Concepts: Core Readings*. Cambridge, Mass.: A Bradford Book, 1999. p. 189–206.
- ROSCH, E.; MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, v. 7, n. 4, p. 573–605, out. 1975.
- 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, abr. 1999.
- RUSE, M.; WILSON, E. O. Moral Philosophy as Applied Science. *Philosophy*, v. 61, n. 236, p. 173–192, 1986.
- SAMET, J. The Historical Controversies Surrounding Innateness. In: E. N. Zalta (Org.); *The Stanford Encyclopedia of Philosophy*. Fall 2008 ed. Disponível em: <<http://plato.stanford.edu/archives/fall2008/entries/innateness-history/>>. Acesso em: 26/11/2016.
- SAMUELS, R. Innateness in cognitive science. *Trends in Cognitive Sciences*, v. 8, n. 3, p. 136–141, mar. 2004.
- SANFEY, A. G. *et al.* The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science*, v. 300, n. 5626, p. 1755–1758, 13 jun. 2003.
- SAYRE-MCCORD, G. Metaethics. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Summer 2014 ed. Disponível em: <<http://plato.stanford.edu/archives/sum2014/entries/metaethics/>>. Acesso em: 20 jul. 2016.
- SCHALLER, M.; PARK, J. H. The Behavioral Immune System (and Why It Matters). *Current Directions in Psychological Science*, v. 20, n. 2, p. 99–103, 1 abr. 2011.
- SCHNALL, S. *et al.* Disgust as embodied moral judgment. *Personality and social psychology bulletin*, v. 34, n. 8, p. 1096–1109, 2008.
- SCHWARTZ, S. H.; BILSKY, W. Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, v. 58, n. 5, p. 878–891, 1990.
- SEARLE, J. R. How to Derive “Ought” From “Is”. *Philosophical Review*, v. 73, n. 1, p. 43–58, 1964.
- SHERIF, M. *et al.* *The Robbers Cave Experiment: Intergroup Conflict and Cooperation*. University Book Exchange, 1961.
- SHWEDER, R. A.; MAHAPATRA, M.; MILLER, J. Culture and Moral Development. In: KAGAN, J.; LAMB, S. (Org.). *The Emergence of Morality in Young Children*. 1. ed. Chicago: University Of Chicago Press, 1987. p. 1–83.
- SIMNER, M. L. Newborn’s response to the cry of another infant. *Developmental Psychology*, v. 5, n. 1, p. 136–150, 1971.
- SINCLAIR, S. *et al.* Social tuning of automatic racial attitudes: the role of affiliative motivation. *Journal of Personality and Social Psychology*, v. 89, n. 4, p. 583–592, out. 2005.
- SINGER, P. *The Life You Can Save: How to Do Your Part to End World Poverty*. Reprint ed. New York: Random House Trade Paperbacks, 2010.

- SINGER, T. *et al.* Empathic neural responses are modulated by the perceived fairness of others. *Nature*, v. 439, n. 7075, p. 466–469, 26 jan. 2006.
- SINNOTT-ARMSTRONG, W.; YOUNG, L.; CUSHMAN, F. Moral intuitions as heuristics. In: DORIS, J. M. (Org.). *The moral psychology handbook*. 1. ed. Oxford; New York: Oxford University Press, 2010. p. 246–272.
- SKORUPSKI, J. The Definition of Morality. *Royal Institute of Philosophy Supplements*, v. 35, p. 121–144, set. 1993.
- SLOVIC, P. *et al.* The affect heuristic. *European Journal of Operational Research*, v. 177, n. 3, p. 1333–1352, 16 mar. 2007.
- SMITH, A. *The Theory of Moral Sentiments*. Excercere Cerebrum Publications, 2014.
- SMITH, E. E. Concepts and Categorization. In: SMITH, E. E.; OSHERSON, D. N. (Org.). *Invitation to Cognitive Science*. Cambridge, Mass.: MIT Press, 1995. p. 1–29.
- SMITH, M. *The Moral Problem*. 1. ed. Oxford, UK ; Cambridge, Mass., USA: Wiley-Blackwell, 1994.
- SOBER, E. Innate knowledge. In: CRAIG, E. (Org.). *Routledge Encyclopedia of Philosophy*. London; New York: Routledge, 1998. v. 4. p. 794–7.
- SOBER, E. *Philosophy Of Biology*. 2. ed. Boulder, Colo: Westview Press, 2000.
- SOBER, E.; WILSON, D. S. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. New ed. Cambridge, Mass.: Harvard University Press, 1999.
- SONG, M.; SMETANA, J. G.; KIM, S. Y. Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology*, v. 23, n. 4, p. 577–582, 1987.
- SPELKE, E. S. The origins of physical knowledge. In: WEISKRANTZ, L. (Org.). *Thought Without Language*. Oxford; New York: Oxford University Press, 1988. p. 168–184.
- SPENCE, M. J.; DECASPER, A. J. Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development*, v. 10, n. 2, p. 133–142, 1 abr. 1987.
- SPERBER, D. Anthropology and Psychology: Towards an Epidemiology of Representations. *Man*, v. 20, n. 1, p. 73–89, 1985.
- SPERBER, D. *Explaining Culture: A Naturalistic Approach*. 1. ed. Oxford, UK ; Cambridge, Mass: Blackwell Publishers, 1996.
- SPERBER, D. In Defense of Massive Modularity. In: DUPOUX, E. (Org.). *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*. Cambridge, Mass.: MIT Press, 2002. v. 47. p. 47–58.
- SPERBER, D. Modularity and Relevance: How Can a Massively Modular Mind Be Flexible and Context-Sensitive? In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind: Structure and Contents*. New York: Oxford University Press, 2005. p. 53–68.
- SPERBER, D.; CLAUDIÈRE, N. Defining and explaining culture (comments on Richerson and Boyd, Not by genes alone). *Biology & Philosophy*, v. 23, n. 2, p. 283–292, 25 maio 2006.
- SRIPADA, C. Nativism and moral psychology: Three models of the innate structure that shapes the contents of moral norms. In: SINNOTT-ARMSTRONG, W. (Org.). *Moral psychology, The Evolution of Morality: Adaptations and Innateness*. Cambridge: MIT Press, 2008. v. 1. p. 319–343.
- SRIPADA, C. S.; STICH, S. A Framework for the Psychology of Norms. In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind: Culture and Cognition*. 1. ed. New York: Oxford University Press, 2007. v. 2. p. 280–301.
- SRIVASTAVA, A. Mean, Dangerous, and Uncontrollable Beasts: Mediaeval Animal Trials. *Mosaic (Winnipeg)*, v. 40, n. 1, p. 127, 1 mar. 2007.
- STERELNY, K. Moral nativism: A sceptical response. *Mind & Language*, v. 25, n. 3, p. 279–297, 2010.
- STERELNY, K. *The evolved apprentice: how evolution made humans unique*. Cambridge, Mass.: MIT Press, 2012.

- STERELNY, K. *Thought in a Hostile World: The Evolution of Human Cognition*. 1. ed. Malden, MA: Wiley-Blackwell, 2003.
- STICH, S. Moral philosophy and mental representation. In: STICH, S. *Collected Papers, Volume 2: Knowledge, Rationality, and Morality, 1978-2010*. New York: OUP USA, 2012. .
- STICH, S. Replies. In: MURPHY, D.; BISHOP, M. (Org.). *Stich and His Critics*. 1. ed. Chichester, U.K. ; Malden, MA: Wiley-Blackwell, 2009. .
- STOTZ, K. Human nature and cognitive–developmental niche construction. *Phenomenology and the Cognitive Sciences*, v. 9, n. 4, p. 483–501, 12 out. 2010.
- SUHLER, C. L.; CHURCHLAND, P. Can Innate, Modular “Foundations” Explain Morality? Challenges for Haidt’s Moral Foundations Theory. *Journal of cognitive neuroscience*, v. 23, n. 9, p. 2103–2116–2122, 2011.
- TAJFEL, H. *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge; New York: Cambridge University Press, 1981.
- TAJFEL, H. *et al.* Social categorization and intergroup behaviour. *European Journal of Social Psychology*, v. 1, n. 2, p. 149–178, 1 abr. 1971.
- TAYLOR, P. W. On Taking the Moral Point of View. *Midwest Studies In Philosophy*, v. 3, n. 1, p. 35–61, 1 set. 1978.
- TIMMONS, M. *Morality without Foundations: A Defense of Ethical Contextualism*. 1 edition ed. New York: Oxford University Press, 1999.
- TISAK, M. S.; TURIEL, E. Children’s Conceptions of Moral and Prudential Rules. *Child Development*, v. 55, n. 3, p. 1030–1039, 1984.
- TOOBY, J.; COSMIDES, L.; BARRETT, H. C. Resolving the Debate on Innate Ideas: Learnability Constraints and the Evolved Interpenetration of Motivational and Conceptual Functions. In: CARRUTHERS, P.; LAURENCE, S.; STICH, S. (Org.). *The Innate Mind: Structure and Contents*. New York: Oxford University Press, 2005. p. 305–337.
- TURIEL, E. *The Development of Social Knowledge: Morality and Convention*. Cambridge; New York: Cambridge University Press, 1983.
- VALDESOLO, P.; DESTENO, D. Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science*, v. 17, n. 6, p. 476–477, 1 jun. 2006.
- VAN ROOJEN, M. Moral Cognitivism vs. Non-Cognitivism. In: ZALTA, E. N. (Org.). *The Stanford Encyclopedia of Philosophy*. Fall 2015 ed. Disponível em: <<http://plato.stanford.edu/archives/fall2015/entries/moral-cognitivism/>>. Acesso em: 20 jul. 2016.
- WADDINGTON, C. H. Canalization of Development and the Inheritance of Acquired Characters. *Nature*, v. 150, n. 3811, p. 563–565, 1 nov. 1942.
- WADDINGTON, C. H. *The evolution of an evolutionist*. Ithaca, NY: Cornell University Press, 1975.
- WAHL, K. *Wie kommt die Moral in den Kopf?: Von der Werteeziehung zur Persönlichkeitsförderung*. Berlin: Springer Spektrum, 2014.
- WALLACE, G.; WALKER, A. D. M. *Definition of Morality*. London: Methuen young books, 1970.
- WARNEKEN, F.; TOMASELLO, M. Altruistic Helping in Human Infants and Young Chimpanzees. *Science*, v. 311, n. 5765, p. 1301–1303, 3 mar. 2006.
- WESTERMARCK, E. *The History of Human Marriage*. London: BiblioLife, 2009.
- WHEATLEY, T.; HAIDT, J. Hypnotic disgust makes moral judgments more severe. *Psychological Science*, v. 16, n. 10, p. 780–784, 2005.
- WHITELEY, C. H. On Defining “Moral”. *Analysis*, v. 20, n. 6, p. 141–144, 1960.
- WILLIAMS, B. *Ethics and the Limits of Philosophy*. 1. ed. London: Routledge, 2006.
- WILLIAMS, B. *Morality: An Introduction to Ethics*. Reprint ed. Cambridge ; New York: Cambridge University Press, 1993.

- WILLIAMS, G. C. *Adaptation and Natural Selection*. Reprint ed. Princeton University Press, 1996.
- WITTGENSTEIN, L. *Philosophical Investigations*. Tradução G. E. M. Anscombe. Oxford: Blackwell, 1986.
- YAU, J.; SMETANA, J. G. Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, v. 74, n. 3, p. 647–658, jun. 2003.
- ZAJONC, R. B. Feeling and thinking: Preferences need no inferences. *American Psychologist*, v. 35, n. 2, p. 151–175, 1980.