

# Moral Nativism: Some Controversies

ROGER V.V. REX *University of Brasília*  
PAULO C. ABRANTES *University of Brasília*

*ABSTRACT:* This paper scrutinizes two research programs that advocate respectively for the existence of a universal moral grammar and a predisposition to moralize behaviours with certain contents. It focuses on how the arguments commonly used to ground each program fare at relevant contemporary research in cognitive science and how well they meet constructivist arguments proposed by Jesse Prinz and Kim Sterelny, among others. We argue that there is little evidence that our moral judgements follow the model of principles and parameters. At the same time, 'ease of learning' suggests that the human brain is somehow prepared to learn moral rules.

*RÉSUMÉ :* Cet article analyse deux programmes de recherche qui soutiennent, respectivement, l'existence d'une grammaire morale universelle et d'une prédisposition à assigner des valeurs morales à des comportements liés à certains contenus. Nous proposons d'évaluer les arguments utilisés par chaque programme en regard de la recherche contemporaine en sciences cognitives, et aussi de vérifier s'ils répondent aux objections constructivistes proposées entre autres par Jesse Prinz et Kim Sterelny. Nous montrons qu'il n'y a pas de preuves suffisantes comme quoi nos jugements moraux suivent le modèle des principes et des paramètres. La «facilité d'apprentissage» suggère, néanmoins, que le cerveau humain est en quelque sorte préparé à apprendre les règles morales.

**Keywords:** moral nativism, innateness, universal moral grammar, moral foundations theory, affective resonance, Darwinism

*Dialogue* 56 (2017), 21–44.

© Canadian Philosophical Association/Association canadienne de philosophie 2016

doi:10.1017/S001221731600072X



## 1 Universal Moral Grammar<sup>1</sup>

The most discussed and well-known version of moral nativism employs language as a model to explain moral cognition, conveying the idea that there is a universal moral grammar (UMG).<sup>2</sup> According to this theory, the morality of each individual is the result of a combination between a universal set of principles and parameter settings emerging from cultural interactions.<sup>3</sup> The main arguments used to support the existence of a UMG focus on *the poverty of the stimulus*. According to this argument, there is a problem in explaining a cognitive ability when there is a gap between the complexity of the learning object and the resources available to the learner. The existence of this gap leads us to posit some kind of innate structure responsible for the learning ability.<sup>4</sup>

Defenders of the UMG theory argue that the behaviour and the moral discourse of adults, to whom children are exposed, would not convey sufficient information to allow the acquisition of a moral competence unless there were learning mechanisms dedicated specifically to the moral domain.<sup>5</sup> The first difficulty with this argument is the fact that moral principles are, as a rule, made explicit. This can be compared to language acquisition: in this domain, children do not receive a significant amount of explicit negative information because different languages are full of enigmatic principles to which the speakers don't have conscious access. In general, ungrammatical sequences are not produced and labelled as deviant; they are just not produced. With respect to morality, however, forbidden acts are produced and described, and the prohibitions are made explicit by means of intensive instruction.<sup>6</sup>

But Marc D. Hauser and John M. Mikhail sustain that moral judgements are also mediated by principles inaccessible to our consciousness. They argue for

---

<sup>1</sup> We believe that the cognitivist debate analyzed in this paper impinges deeply on moral philosophy, although we admit that many moral philosophers still think otherwise. They believe that morality deals with what we ought to do, while cognitive sciences just describe how things are. Thus, research in this area would be worthless for moral philosophy. Nonetheless, we argue that empirical studies about morality give us a kind of 'picture' of the values prevailing in different human groups, a comprehension of the origins of these values, and hence predictions about the possibilities and ways of changing them. Any normative theory that seeks some practical application should take these factors into account. There is no point, for instance, in recommending behaviours that people are incapable of following, i.e., 'ought' implies 'can.' For a defence of this position, see Ruse and Wilson (1986); and Mikhail (2011, Chapter 7).

<sup>2</sup> Dwyer, 2007; Hauser, 2007; Mikhail 2011; Hauser, Young and Cushman, 2008.

<sup>3</sup> Hauser et al., 2008, p. 122.

<sup>4</sup> Sripada, 2008, p. 325–326.

<sup>5</sup> Prinz, 2008, p. 430; Joyce, 2006, p. 137.

<sup>6</sup> Sterelny, 2010, p. 290–291.

this thesis based on the evidence from *trolley problems*.<sup>7</sup> According to Hauser and Mikhail, the respondents apply the principle of double effect<sup>8</sup> to find answers to these dilemmas, although they are unable to formulate the principle. Since this principle is rarely articulated and it is not an obvious generalization from particular cases, it is hard to conceive how it could be learned. If people do not know it consciously, they are unable to teach children about it. If children nevertheless learn it, they seem to be somehow prepared for this.<sup>9</sup> The alleged solution is that principles such as these are inscribed on each individual as part of a UMG.<sup>10</sup>

Constructivists, by contrast, do not interpret the decisions taken by the respondents to those problems as the application of unconscious rules or principles specific to morality. Jesse Prinz, for example, believes instead that people learn that killing is wrong through paradigmatic cases and they are more tolerant of behaviours that deviate from those cases.<sup>11</sup> The amazing convergence of answers to the classic trolley problem<sup>12</sup> could be explained by the fact that the solution considered morally forbidden—pushing the fat person onto the tracks—combines a set of characteristics normally considered reprehensible in different societies: the fat person is being employed as a means to an end, the case is much like the paradigmatic cases of murder, and there is a strong personal element.<sup>13</sup> The act of pushing someone is much more personal than the act of pulling a lever, which encourages more people to consider the pushing impermissible. When these factors are isolated, the participants fail to converge towards a common response. This fact becomes salient when we analyze the results of research involving two scenarios called “Loop Track” and “Man-in-Front” by Mikhail.<sup>14</sup> The essential difference between the two scenarios is that in the first one the death of the person located on the adjacent rail is a necessary

---

<sup>7</sup> Hauser et al., 2008; Mikhail, 2011, p. 38.

<sup>8</sup> This is how Gilbert Harman describes the principle of double effect: “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” (2008, p. 346).

<sup>9</sup> Roedder and Harman, 2010, p. 286–287.

<sup>10</sup> Harman, 2000, p. 225.

<sup>11</sup> Prinz, 2013, p. 106.

<sup>12</sup> Regarding the classic version of the dilemma, originally formulated by Philippa Foot, the vast majority of respondents replied it was morally permissible to divert the trolley to an adjacent rail to save the lives of five workers, even if that conduct caused the death of an innocent. However, they did not consider it permissible to push a person in front of the trolley in order to save those same five workers. Just a few respondents could explain why only one of these behaviours was permissible.

<sup>13</sup> For the distinction between personal and impersonal acts, see Greene, 2001.

<sup>14</sup> Mikhail, 2011, p. 107–108.

means to prevent the deaths of five others; while in the second case, the death of that person is just a side effect of diverting the train. In the first scenario, 48 percent of respondents considered diverting the train permissible; in the second situation, 62 percent felt this way. Although the answers diverged greatly, Mikhail believes that the principle of double effect was applied since the majority of participants judged impermissible the act that caused injury as a means for a greater good.

In spite of this, the question that seems important here is the extent to which we can, from such divergent results, identify universal principles that purportedly have produced the moral judgements observed. The fact that it is possible to come up with principles such as the double effect to make sense of the moral intuitions of most respondents does not mean that their intuitions have been produced by means of the alleged principles. Therefore, it is important to distinguish between an external and an internal approach. The external approach seeks to develop a set of principles that explain most of the intuitions expressed by the participants, and it is always possible to elaborate more than one set of such principles that can fulfil the explanatory task equally well. The internal approach, on the other hand, seeks to identify the principles that are causally responsible for the observed intuitions. Shaun Nichols points out that UMG supporters use the trolley problems to come up with a set of principles that is consistent with most intuitions without bothering to show that those principles are actually involved in the causal production of the intuitions.<sup>15</sup>

The trouble that the intuitions of many people are not compatible with the supposed universal principles is circumvented by appealing to the distinction between competence and performance.<sup>16</sup> Mikhail and Hauser believe that certain psychological limitations, performance errors, emotions, and other factors distort moral judgements. According to them, the solution is to isolate those factors and to identify moral judgements that exhibit the optimal operation of the moral faculty.<sup>17</sup> However, it is unclear whether what they consider interference, especially certain emotions, is not in fact constitutive of our morality.

There is a difficulty in combining the idea that moral judgements conform to a grammatical structure and the idea that emotions have a causal role in the production of intuitions. Mikhail, Hauser, and Susan Dwyer, while recognizing the existence of an important interface between cognition and emotion, prefer to argue that moral judgements cause emotions and not vice versa. Confronted with empirical evidence that we can modify moral judgements through the manipulation of emotions by means of hypnosis and environmental changes,<sup>18</sup> they seek to preserve their theories by appealing, again, to the distinction

---

<sup>15</sup> Nichols, 2005, p. 360.

<sup>16</sup> For the competence and performance distinction, see Chomsky (1965, p. 3–4).

<sup>17</sup> Mikhail, 2011, p. 103.

<sup>18</sup> Wheatley and Haidt, 2005; Schnall et al., 2008.

between competence and performance: they argue that emotions just affect moral performance, but are not constitutive of the moral faculty itself. Emotional processes sparked before the purported computational processes required to produce moral judgements can affect the latter and motivate certain actions, but are not necessary for the functioning of the moral faculty.<sup>19</sup> There is, however, evidence that emotional responses are required for normal moral development. The lack of empathy in children with psychopathic tendencies, for example, jeopardizes this development.<sup>20</sup>

Yet, in order to ward off the idea that we have conscious access to the reasons underlying our moral judgements, some advocates of UMG argue that moral justifications are not connected with moral judgements. Dwyer, for instance, points to the phenomenon described by Jonathan Haidt as “moral dumbfounding”:<sup>21</sup> under certain circumstances—e.g., when considering dilemmas about incest—people simply cannot find reasons to justify their moral judgements. Defenders of UMG believe that this kind of phenomenon occurs because the principles that led to the decision of these people (operative principles) are inaccessible to consciousness.<sup>22</sup> According to Haidt, we overestimate the role that conscious reflection plays in the forming of our moral judgements. In most cases, this conscious process serves only to rationalize intuitive judgements. However, Haidt points out that reflection is not just an epiphenomenon of morality. In some circumstances, it can modify over time the intuitive judgement of the individual, and in many cases it serves to change the intuitions of other group members.<sup>23</sup> The disgust vegetarians have for meat, for instance, is a consequence, not a cause, of their moral convictions.<sup>24</sup> Therefore, there might be, in the context of morality, a *diachronic interaction* between intuitive reactions on the one side, and conscious and articulated thought on the other side. Explaining the interaction between these two kinds of cognition is a major challenge faced by UMG theorists, and they do not have a convincing model to propose. The analogy between language and morality cannot help here, because the processing of syntactic rules does not rely in any way on conscious reasoning.<sup>25</sup>

Hauser, Dwyer, and Bryce Huebner, even when they recognize that children receive explicit instructions about morality, argue that they are not sufficient to explain the sophisticated moral rules found in their judgements. In particular, the imperatives they hear—‘always keep your promises’—would not be enough

<sup>19</sup> Dwyer, Huebner and Hauser, 2010, p. 495.

<sup>20</sup> Blair et al., 2006, 1997; Dupoux and Jacob, 2007, p. 376.

<sup>21</sup> Dwyer, 2009; Haidt, 2001, p. 817.

<sup>22</sup> Mikhail, 2011, p. 83; Dupoux and Jacob, 2007, p. 374.

<sup>23</sup> Haidt and Bjorklund, 2008.

<sup>24</sup> Fessler et al., 2003.

<sup>25</sup> Sterelny, 2010, p. 282–284.

to explain how children are able to identify exceptions to these rules.<sup>26</sup> Still, the explicit instructions are not as rough as Hauser, Dwyer, and Huebner believe. Parents usually offer explanations for their instructions. They resort to different reasons and norms to justify why the child should change her behaviour. In many occasions, parents seek to induce her to adopt the perspective of the offended person, indicating the suffering caused and making it clear that the child was responsible for that harm.<sup>27</sup>

In addition to explicit instructions, children receive other stimuli. Stories, myths, and children's songs are full of moral lessons. Moreover, children's lives are not passive. They do not just listen to stories and watch what others do. They act and face conflicts with other children involving harm and distribution of resources. Perhaps children would not be able to learn the rules for distribution of goods just by observing the behaviour of their close relatives; but they often take part in discussions in which they seek to find suitable rules for resolving these types of conflict—e.g., '*I divide and you choose*.' Interactions such as these are part of the learning process and are loaded with information related to morality.<sup>28</sup>

In short, it is not difficult to show that explicit and implicit moral stimuli are plentiful in the environment in which children develop. However, this is not sufficient to refute the poverty of the stimulus argument. The refutation requires a demonstration that the information received is rich enough to explain the capacity developed, since the poverty of the stimulus is not only related to the amount of stimuli but also to the complexity of what is learned.

Concerning morality, the subject to be learned seems in a way to be much simpler than in the case of language. Moral norms are not as obscure as the recursive and structural rules of grammar. The development of a moral capacity requires learning a set of more specific rules, for example, 'share your toys,' 'do not hit other children,' 'respect your elders,' etc.<sup>29</sup>

Notwithstanding, defenders of UMG employ another argument to support the complexity of the task of making moral judgements: how can we explain the development of a capacity to judge an infinite number of cases from a finite experience? Thus, scholars who seek to explain morality would also face, like the linguists, the *projection problem*, i.e., the problem of explaining how individuals are able to apply their moral knowledge to cases different from those they experienced previously.<sup>30</sup> A mature speaker of any language has been in touch with only a limited number of sentences throughout her life. Despite this, she is able to build an infinite number of sentences that are comprehensible to

---

<sup>26</sup> Dwyer, Huebner and Hauser, 2010, p. 492.

<sup>27</sup> Hoffman, 2001, p. 143.

<sup>28</sup> Sterelny, 2010, p. 288–291.

<sup>29</sup> Sripada, 2008, p. 328.

<sup>30</sup> Mikhail, 2011, p. 30.

other speakers. Given the limited storage capacity of our brains, it is impossible that all these sentences are stored individually. So, there must be a grammar by which we can, from a limited vocabulary and sentence patterns, build an unlimited number of expressions.<sup>31</sup>

UMG enthusiasts propose that something similar takes place in the realm of morality: an individual with a developed sense of justice would be able to make an unlimited number of intuitive judgements about the moral properties of actions and agents.<sup>32</sup> Again, considering the limited storage capacity of the brain and the unlimited ability to make judgements, it follows that the moral faculty is more than a simple list. There would be a cognitive system composed of principles and rules responsible for producing an unlimited number of moral judgements.<sup>33</sup>

There is, however, a big difference between morality and language: to explain how we are able to interpret completely new phrases from the sounds we hear, we must appeal to a complex set of principles, since it is hard to explain the reversibility in language. How is it possible that someone produces a new set of sounds from semantic representations and someone else reconstructs these semantic representations based on that combination of sounds never heard before? When a native speaker makes a judgement about the grammaticality of a sentence, she does not judge only if that sentence is correct, she also seeks to understand (rebuild) what the other person meant. In moral judgements, this kind of reversibility is absent. While language is the result of a generative system, morality is simply the outcome of an evaluative system.<sup>34</sup> This means that morality looks more like our ability to evaluate the taste of a food, or the quality of a work of art, than our capacity for language.

In this context, constructivist theories based on the ability of pattern recognition have the advantage of simplicity over UMG, as they explain moral learning through mechanisms employed in other forms of knowledge. The first step to postulate the existence of a specific moral learning capacity would be to demonstrate the impossibility of learning to make moral judgements through general-purpose learning systems, such as pattern recognition. No one doubts that we have this ability. Thus, before proposing the existence of a special-purpose system, we should explain why general-purpose learning systems are unable to provide an adequate explanation.<sup>35</sup>

The human mind is good at recognizing patterns and noticing similarities between different situations. The exercise of this capacity results in intuitive judgements about new cases. Pattern recognition can be extended to infinite

---

<sup>31</sup> Mikhail, 2011, p. 45–46; Dwyer, Huebner and Hauser, 2010, p. 488–489.

<sup>32</sup> Dwyer, Huebner and Hauser, 2010, p. 489.

<sup>33</sup> Mikhail, 2011, p. 46 and p. 72.

<sup>34</sup> Dupoux and Jacob, 2007, p. 376.

<sup>35</sup> Prinz, 2012, p. 128 and 150.

new cases and takes place rapidly and automatically. In addition, the person who recognizes the pattern often cannot explain which features motivated the recognition. An expert on birds, for example, can identify at a glance certain species, even when she is unable to explain exactly how she did this. So, if moral intuitions were the result of a pattern recognition process, the fact that they are fast and that we have difficulty in distinguishing their causes, would not be a surprise.<sup>36</sup>

Kim Sterelny takes the relationship between tacit and explicit principles present in morality as something very similar to what happens with other skills acquired through a general-purpose learning system. An artisan, for example, has considerable explicit knowledge—she can easily explain many of the techniques she uses—which coexists with implicit knowledge resulting from the habit of practicing her profession and with a capacity to recognize patterns intuitively.<sup>37</sup> However, this distinction between explicit knowledge and tacit knowledge, both in the moral context and in relation to artisan activity, is not clear cut: the artisan may be able to explain why she chose or rejected a certain material, but this reconstruction ends up being partial because she does not have full access to the reasons that really motivated her choice, which means that explicit and implicit knowledge are interwoven.<sup>38</sup>

If moral judgements can be explained by pattern recognition processes, the main arguments of UMG proponents collapse, including the poverty of the stimulus, and the solution they propose for the projection problem. They must show that the subtleties and abstractions involved in moral judgements make the pattern recognition explanation implausible. So far, efforts employed by Hauser and Mikhail to exemplify the subtleties involved in moral judgements have focused on issues involved in our abilities to deal with the complexity of the social environment, such as our ability to attribute mental states to others and to mentally represent the characteristics of an action. But these abilities are not specific to morality.<sup>39</sup> Their general character makes a case against UMG's claim that we have a specific moral faculty.

## **2 Moral Foundations and Affective Resonance**

Morality exists in all human societies we know of and almost every individual develops a sense of it without formal instruction and without intentional effort.<sup>40</sup> Haidt elaborated his Moral Foundations Theory (MFT) based on these universal features. Together with Craig Joseph,<sup>41</sup> he identified common principles of

---

<sup>36</sup> Sterelny, 2010, p. 287–289.

<sup>37</sup> Sterelny, 2012.

<sup>38</sup> Sterelny, 2010, p. 293–294.

<sup>39</sup> Sterelny, 2012; Sterelny, 2010, p. 287–288.

<sup>40</sup> Joyce, 2006; Ayala, 2010, p. 9016.

<sup>41</sup> Haidt and Joseph, 2004.



human morality through an analysis of five studies on universal characteristics.<sup>42</sup> After trying to list all things humans and chimpanzees value in the behaviour of others, they summed up their findings in five categories: (a) sensitivity or aversion to pain signals and suffering in others (harm/care); (b) negative responses to those who fail to reciprocate favours (fairness/reciprocity); (c) anger against those who fail to show signs of deference and respect (authority/respect); (d) emotions related to disgust, required to explain moral rules about food, sex, menstruation, and disposal of corpses (purity/sanctity); (e) attitudes towards group boundaries (in-group/out-group).<sup>43</sup> Haidt claims that these five categories of intuitions comprise the moral foundations.<sup>44</sup> He argues that each one is associated with a mental module and is connected to different families of emotions: suffering leads to empathy and compassion; disregard for hierarchy produces resentment and contempt; violations of reciprocity generate anger and guilt; violations related to purity provoke disgust. These domains are supposed to be universal, but each culture can define their specific contents.<sup>45</sup>

Haidt wondered if our capacity to make moral judgements relied on something similar to taste receptors. Cuisines are cultural products and each one is unique and has a set of key ingredients. Even so, they are constructed on the basis of a sensory system that includes only five types of taste receptors. The structures of our tongues, noses, and brains restrict the number of possible types of cuisines, but leave plenty of space to creativity.<sup>46</sup> Haidt's thesis is that there is something similar in morality. That is, although the moral matrix of a given group is a cultural construct influenced by various particular historical and environmental factors, it must be compatible with minds equipped with certain types of social receptors.<sup>47</sup>

According to Haidt, the clustering of moral rules around certain themes is the result of innate predispositions that facilitate the learning of certain norms. In psychology, it is universally accepted that some things are easier to learn than others. It is extremely difficult to mould a child's mind when the effort is made in the opposite direction to what she likes naturally. It does not require much effort, for example, to make a child prefer candies to broccoli, the sympathy of other children rather than the approval of adults, or to retaliate against aggression instead of loving her enemies.<sup>48</sup> The central idea of this kind of model is that humans have strong predispositions to develop certain reactions

---

<sup>42</sup> Brown, 1991; Fiske, 1992; Schwartz and Bilsky, 1990; Shweder et al., 1997; de Waal, 1996.

<sup>43</sup> In some other works, Haidt calls this domain "Loyalty/Betrayal Foundation" (Haidt, 2012).

<sup>44</sup> Haidt and Bjorklund, 2008, p. 202–203; Haidt and Kesebir, 2010, p. 822.

<sup>45</sup> Haidt and Joseph, 2004.

<sup>46</sup> Haidt and Bjorklund, 2008, p. 201–202.

<sup>47</sup> Haidt, 2012, Chapter 6.

<sup>48</sup> Haidt and Bjorklund, 2008, p. 201.

and preferences.<sup>49</sup> What we consider morally permissible, forbidden, or obligatory may be a result of these innate emotional propensities combined with our cultural experiences, without the need to apply a complex set of principles, such as those involved in language.

The hypothesis that human beings have a predisposition (preparedness) to moral learning is reinforced by experiments that show the early development of perceptions that are involved in moral judgements. Before walking, children are already able to recognize and value behaviours, such as helping or harming others. Kiley J. Hamlin, Karen Wynn, and Paul Bloom presented to children aged between 6 and 10 months some performances in which a puppet was trying to climb a slope. In some of the presentations, the puppet was helped by another puppet that pushed it up. In others, a third puppet appeared on the slope and struck the rising puppet, preventing it from reaching the top. After these presentations, both puppets were put in front of the children, who showed a strong preference for the one who tried to help.<sup>50</sup> According to Haidt and Serlin Kesebir, experiments like this indicate the existence of a perception system capable of creating positive emotions towards helpers, and negative emotions towards bullies.<sup>51</sup> Emotions like these, in turn, could facilitate the adoption of rules designed to avoid actions that harm third parties.

In turn, this emphasis on the role of emotions brings us to the *affective resonance* model elaborated by Nichols. Inspired by Dan Sperber, he advocates an epidemiological model that emphasizes how emotional dispositions can restrict moral development possibilities. The hypothesis formulated by Nichols predicts that, all other circumstances kept unchanged, rules prohibiting actions that have a high probability of awakening negative emotions have a higher chance of being assimilated and transmitted than rules that are unconnected to emotions.<sup>52</sup> From Nichols' perspective, our emotional dispositions influence our moral judgements, but should not be confused with them.<sup>53</sup> He tries to explain how emotions and norms interact. To this end, he resorts to the major role cultural evolution played in determining which norms have 'survived' throughout history. Those rules prohibiting actions that are likely to produce negative emotions have more cultural fitness.<sup>54</sup> He acknowledges that there

---

<sup>49</sup> Giroux, 2011, p. 292–293.

<sup>50</sup> Hamlin, Wynn and Bloom, 2007.

<sup>51</sup> Haidt and Kesebir, 2010, p. 804.

<sup>52</sup> Nichols, 2008, p. 270.

<sup>53</sup> For this reason, experiences with children who show aversion to harmful practices should be interpreted with caution. The fact that a baby expresses aversion to certain offences does not mean that she judges that this behaviour is morally wrong.

<sup>54</sup> Affective resonance hypothesis: "Norms that prohibit actions to which we are predisposed to be emotionally averse will enjoy enhanced cultural fitness over other norms" (Nichols, 2008, p. 269–270).

are numerous factors that influence cultural evolution and affective resonance is just one of those: emotionally conspicuous cultural aspects tend to attract our attention and to be memorized, and are, therefore, more likely to endure.<sup>55</sup>

According to Nichols' hypothesis, rules designed to prevent the production of damage and injury (harm norms) have an advantage over other rules in the process of cultural evolution, since 'normal' human beings have a strong aversive reaction to suffering. As well as other basic emotions (such as sadness, anger, disgust, and fear), emotional reactions to the suffering of others are believed to be universal and innately specified. As a result, in all cultures, actions that cause suffering to others tend to arouse aversion.<sup>56</sup> Rules to prevent harm could have emerged for different reasons at different times, but the important point is that when this happens they find an important ally in emotions.

Thus, Nichols explains the existence of certain moral norms as a result of innate biases.<sup>57</sup> Yet, he stresses that these biases are part of an affective innate system, which means they are neither information in a propositional form nor a set of innate moral principles.<sup>58</sup> For Nichols, the moral domains originate from emotional systems, especially the affective system that responds to suffering in others. If this type of model is correct, emotions can play a role in shaping cognitive structures, reducing the appeal of theories that propose the existence of innate propositional knowledge.<sup>59</sup>

Although Nichols emphasizes Haidt's harm/care foundation, he acknowledges the existence of other candidates for the role of moral universals. He admits the importance of emotions associated with fairness and, to a lesser extent, disgust in shaping morality.<sup>60</sup> The history of etiquette norms, for example, demonstrates that norms associated with the emotion of disgust have a greater chance of remaining unchanged.<sup>61</sup>

---

<sup>55</sup> Nichols, 2008, p. 269–270.

<sup>56</sup> Nichols, 2008, p. 271.

<sup>57</sup> Sripada offers a similar explanation. He argues that changes in moral norms would be better explained by what he calls "thematic clustering." In virtually all human groups, morality deals with some common issues. Sripada says that the existence of this thematic clustering is the result of innate biases acting on the content of moral norms. The central idea of this innate biases model is that there are innate structures favouring the arousal of a feeling of aversion or sympathy towards certain behaviours and, consequently, favouring the emergence and maintenance of certain moral norms (Sripada, 2008, p. 330–337).

<sup>58</sup> Nichols, 2005, p. 369.

<sup>59</sup> Nichols, 2005, p. 368.

<sup>60</sup> Nichols, 2005, p. 356; Nichols, 2008, p. 266.

<sup>61</sup> Nichols, 2002.

The central aspect of the models developed by Haidt and Nichols is the relevance attributed to innate dispositions, especially certain types of emotions, in shaping the morality of a given human group. According to them, there is not an innate moral knowledge, as stated in UMG, but a tendency to moralize certain behaviours as a consequence of the emotions they arouse.

### **3 Shortcomings of the Moral Foundations Approach**

Prinz is one of the main critics of moral nativism in any of the above-mentioned modalities. In general, he attempts to demonstrate that the capacities involved in moral judgements are not adaptations specific to morality. However, this kind of argument is relevant to the refutation of moral nativism only when it presupposes an evolutionary concept of innate character.

Richard Joyce classifies the definitions of innateness in two main groups: the evolutionary conception and the developmental conception. According to the first definition, moral nativism is equivalent to the claim that morality is an adaptation in the Darwinian sense; that is, it was selected for by means of natural selection. On the other hand, those who adopt a developmental conception consider that a trait is innate when its appearance is buffered from variations in the environment where the individual develops.<sup>62</sup>

Haidt explains from an evolutionary standpoint the emergence of emotions related to morality. He asserts that these emotions were selected for, that is, that they are adaptations. Perhaps he believes that the ability to make moral judgements is also an adaptation, but this is not the most important aspect of the kind of moral nativism to which he is committed. The models of Haidt and Nichols hold that innate traits in a developmental sense (either adaptations or side-effects of adaptations) ensure that morality develops and retains certain properties even in different environments. Therefore, Prinz' critiques do not touch the most relevant aspects of the models developed by Haidt and Nichols. Nevertheless, there are two critiques that are more germane to the MFT: the fact that the latter is committed to a modular vision of the human mind and its incompleteness, since it does not provide an explanation for the mental steps previous to the triggering of intuitions.

Haidt adopts in his MFT the idea that the human mind is modular, but rejects some of the features that Jerry A. Fodor<sup>63</sup> uses to characterize a module. He was inspired by Sperber's massive modularity hypothesis:<sup>64</sup> there is a set of innate learning modules capable of producing more specific modules during development.<sup>65</sup> In other words, there is an innate learning module—first-order module—for each moral domain and each of these modules produces, from the

---

<sup>62</sup> Joyce, 2013, p. 532–533.

<sup>63</sup> Fodor, 1983.

<sup>64</sup> Sperber, 2005.

<sup>65</sup> Haidt and Joseph, 2007, p. 379–380.

experiences of the individual, several other working modules responsible for generating moral intuitions in specific situations—second-order modules.<sup>66</sup>

The moral dumbfounding phenomenon is the main reason why Haidt believes the human mind is modular. In many cases, people keep their moral judgements unchanged despite being confronted with new information that invalidates the justifications used to support the original judgements. Haidt thinks the explanation for this is that the systems that produce these judgements are encapsulated to some extent. Unlike Fodor, Haidt admits that the modules have access to information stored elsewhere in the mind, but in a limited way. Haidt also holds that the modules dedicated to morality belong to specific domains.

Nonetheless, it is unclear how Haidt came to the conclusion that each of the five moral foundations corresponds to a different mental module. The simple fact that we are able to classify intuitions about morality in five categories (foundations) does not justify the thesis that each of them is implemented independently or by a discrete computational mechanism. The foundations described by Haidt could be universal and have an innate basis, but that does not mean that there are five specialized modules.<sup>67</sup> While it is true that some things are easier to learn than others, this does not justify the inference that *ease of learning* results from the existence of learning modules dedicated to specific domains.

The idea that the human mind is composed of several specialized mechanisms dedicated to solving particular types of problems faces strong opposition. Even Sperber recognizes that only a small number of cognitive scientists believe that the mind is massively modular. Most of them believe that the mind is largely non-modular. Although many admit the existence of modules related to perception, just a few argue that the central systems that process these inputs are modular.<sup>68</sup> In addition, the idea of modularity proposed by Sperber and adopted by Haidt is not consilient<sup>69</sup> with empirical results from neuroscientific research. If the human mind were made up of modules, we should expect the organization of the brain to reflect the existence of these modules.<sup>70</sup> However, the anatomy of our central nervous system makes the ideas of informational encapsulation and domain specificity implausible.

---

<sup>66</sup> We took this expression—“second-order modules”—from Suhler and Churchland (2011, p. 2104).

<sup>67</sup> Mallon, 2008, p. 151; Giroux, 2011, p. 294–295.

<sup>68</sup> Sperber, 2005, p. 53.

<sup>69</sup> We use the word ‘consilient’ in the sense proposed by Whewell: “The Consilience of Inductions takes place when an Induction, obtained from one class of facts, coincides with an Induction, obtained from another different class. This Consilience is a test of the truth of the Theory in which it occurs” (2012, pos. 7172).

<sup>70</sup> Suhler and Churchland, 2011, p. 2109.

Local neural connections in the cortex are dense, while connections that span greater distances are sparser, a fact known as “small world architecture.” Despite this, just a few synapses separate a particular neuron from any other present in our brains. As stated by Christopher L. Suhler and Patricia Churchland, “everything is easily accessible to everything else in a few synaptic steps.” This pattern prevails even in the primary visual cortex (V1), the area responsible for receiving inputs from the retina through the Lateral Geniculate Nucleus (LGN). Over 80 percent of synaptic contacts present in V1 do not come from LGN, but from other brain regions.<sup>71</sup> As a result, the V1 operation depends on what is going on in these other regions.

The assumption that modules are responsible for specific domains also confronts similar challenges. Prinz demonstrates that the usually mentioned instances of modules do not correspond to specific domains. He points out, for example, how mind-reading relies on working memory.<sup>72</sup> Prinz mentions research conducted by Margaret C. McKinnon and Morris Moscovitch that shows that the performance of individuals whose working memories are kept busy is impaired in tasks related to the attribution of beliefs.<sup>73</sup> Prinz further highlights how neuroimaging studies show that mind-reading involves the use of several brain regions, and that each of those contributes to many other capacities, i.e., they are not specific to any domain. Likewise, moral judgments recruit various areas of the brain responsible for different capacities, including those normally associated with emotional centres.<sup>74</sup>

Haidt is aware of the objections directed towards the thesis of the modularity of mind. As a result, he argues that the existence of modules is not a central aspect of MFT. The foundations he proposes could be explained more generally in terms of preparedness. A milder version of MFT could, therefore, be described as follows: the human mind has been shaped by evolutionary processes in such a way that children learn easily to be concerned with the avoidance of harm, the correction of injustices (fairness), the preservation of the members of their own groups (in-group), the respect for authority, and with purity issues. Notwithstanding, this does not mean that they have any innate moral knowledge; it shows only that they are prepared to acquire certain moral beliefs and to resist others.<sup>75</sup>

When reformulated this way, Haidt’s MFT becomes akin to the epidemiological model proposed by Nichols. According to this model, it is plausible to sustain that the emotional systems tuned to be affected by the suffering of others have evolved as a way of overcoming environmental challenges during

---

<sup>71</sup> Suhler and Churchland, 2011, p. 2109.

<sup>72</sup> Prinz, 2006, p. 28.

<sup>73</sup> McKinnon and Moscovitch, 2007.

<sup>74</sup> Prinz, 2006, p. 29–30.

<sup>75</sup> Haidt and Bjorklund, 2008, p. 204.

the Pleistocene. Furthermore, their influence on cognitive structures is not restricted to specific areas. The emotions that affect moral judgements are not specific to this kind of appraisal, since they can also affect the acquisition of knowledge in other areas. Our responses to suffering could, for instance, affect the way we think about natural disasters that cause human misery.<sup>76</sup> The pervasive influence of emotions on behaviour, inasmuch as it indicates the absence of encapsulation, challenges the idea that the mind has a module or group of modules specific to morality.

Nichols' affective resonance model and what is left of MFT essentially imply that (1) we learn some things more easily than others, and (2) our emotions play an important role in determining what is easier to learn. These statements are not very controversial since it is hard to imagine how dispositions such as our basic emotions—sadness, anger, disgust, fear, etc.—could not interfere with learning and the development of social norms. However, this kind of claim is not enough to clarify many of our questions about moral nativism: Haidt says that the human mind is prepared to learn certain things, but does not provide details about what exactly this innate organization that favours learning comes to be. He does not explain (at the level of cognitive psychology, developmental psychology, neuroscience, etc.) how humans are prepared to acquire moral norms.<sup>77</sup> Similarly, Nichols' claim that our emotions favour the adoption of certain norms needs a more thorough explanation. Haidt and Nichols do not unveil the details of how cognitive processes lead to the production of moral judgements.

Another major critique of MFT relates to the fact that it does not give due attention to cognitive elements prior to the outbreak of emotions and intuitions. A creature can have an emotion only after its mind identifies the situation as worthy of that emotion. There is a previous unconscious analysis responsible for identifying the causes and consequences of an action—who did what? why? by which means? to achieve which goals?—and for triggering an emotional reaction.<sup>78</sup> Hauser and Mikhail point out that very similar actions cause completely different emotional reactions depending on how they are perceived or mentally represented by those who watch them. Depending, for example, on the intentions we attribute to the actor (mind-reading), we will have different reactions. Emotions alone could not explain why an action is considered wrong in one context, but right in another. Therefore, we must understand how mental representations, emotions, and other slower cognitive processes interact. A full explanation of how moral judgements occur should clarify the computational processes responsible for analyzing the scenario where the action to be judged takes place.<sup>79</sup>

---

<sup>76</sup> Nichols, 2005, p. 368.

<sup>77</sup> Suhler and Churchland, 2011, p. 2105.

<sup>78</sup> Hauser, 2007, p. 8.

<sup>79</sup> Hauser, 2007; Mikhail, 2011; Dwyer, Huebner and Hauser, 2010, p. 494.

Moral judgements rely on an interpretation of the action made by the person who judges. When we make these judgements, we use a number of inferences about the intentions and mental states of the actor. The abilities required to make these inferences are not restricted to the realm of morality. We employ them when we interpret any action, no matter how irrelevant it might be to morality. The distinction between intentional and accidental actions, for example, usually is relevant to moral judgements, as many actions are only considered reprehensible when performed intentionally. Notwithstanding, when we observe a child playing with a ball, we are able to distinguish if she let the ball slip from her hands or if she bounced the ball on purpose, although this does not have any moral relevance. Similarly, when we see a child cutting a lemon, we assign a goal to that act, for instance, making lemonade. We constantly infer invisible properties of the mind through indirect verification, for example, what someone else is looking at, which objects this person is trying to reach, or where she came from. A theory of mind is needed to understand the intent and the action of the aforementioned child, but it does not imply a moral judgement. This capacity shares with the capacities for visual perception or memory the fact that they can be used in moral judgements but are not specific to that sphere, i.e., they can also be used in a range of other tasks.<sup>80</sup>

The way we interpret an action can provoke a feeling of aversion or attraction. This feeling motivates the adoption of a certain reaction and explains in large part how we arrive at a moral judgement. However, the simple reaction (aversion or attraction) to a particular circumstance does not depend on the existence of a moral judgement. Even an amoeba is able to identify certain features of its environment—concentration of ions, for instance—and to react, moving closer or moving away. Obviously the amoeba's reaction does not call for the application of a set of principles and parameters with a complexity near to that of a grammar. When we face a particular environment or action we also react according to the characteristics we identify. Nonetheless, the simple emotional reaction (aversion or attraction) to a particular action does not fully explain our moral judgement, since this involves a diverse cognitive element, a belief about the value of the performed action.

The distinction between a simple emotional reaction and a belief is useful to show that morality is not a monolithic entity. That is, it involves a number of elements. As we shall see, realizing this is crucial for answering the fundamental question of the debate on moral nativism.<sup>81</sup>

<sup>80</sup> Sripada, 2008, p. 362; Hauser, 2007, p. 50–51.

<sup>81</sup> Nichols recognizes that morality is not a monolithic entity. He believes that it is a side effect of at least two other features: “both of the mechanisms that I’ve suggested contribute to moral judgment might well be adaptations. However, it is distinctly less plausible that the capacity for core moral judgment itself is an adaptation. It’s more likely that core moral judgment emerges as a kind of by-product of (*inter alia*) the innate affective and innate rule comprehension mechanisms” (2005, p. 369).



#### 4 So, is Morality Innate?

If we adopt a developmental concept of innateness, it will lead us to conclude that morality is innate. After all, it is present in every community we know of and, in general, the existing moral norms in different communities cluster around the same themes. These facts suggest that there is a biological basis for certain biases in moral learning. Even some opponents of moral nativism agree that the acquisition of norms is biologically prepared. However, they believe that the crucial adaptations are perceptual and motivational and not specific to morality. In this context, they are presupposing an evolutionary concept of innateness. If they adopted a developmental concept, they would agree that morality, or at least most of its elements, is innate. Prinz, for example, considers morality a side effect of psychological traits that have evolved for other purposes. He recognizes that morality is constrained by our biological makeup and that we are not born as a *tabula rasa*. According to him, our emotions, the ability to attribute mental states, and the care we have for our relatives function as building blocks of morality, but should not be confused with it.<sup>82</sup>

On the other hand, if we adopt an evolutionary concept of innateness, the issue becomes more complicated. It is likely that many of the aspects related to what Charles Darwin called “social instincts” are adaptations. That is, all those emotions that lead us to cooperate within our societies and that end up producing what Haidt calls ‘moral foundations’ were probably selected for because they favoured certain behaviours. However, morality cannot be reduced to social instincts, as it relies also on what Darwin called “intellectual powers.”<sup>83</sup>

The distinction made by Darwin between *social instincts* and *intellectual powers* is still relevant, since it makes clear that the existence of groups of altruists who do not have a moral faculty is at least conceivable.<sup>84</sup> For animals to behave in one way or another, they do not have to be able to judge any behaviour as good or bad: the natural selection of a behaviour, and consequently its evolution, does not require that it was consciously adopted. That is, there could be some kind of being who behaved in a way we consider morally praiseworthy, but was incapable of making any moral judgement. We could act altruistically by some inclination, without having the belief that we *ought* to act in this way. Doing something because we like it is different from doing something because we believe we have a duty.<sup>85</sup>

Self-consciousness is one of those intellectual capacities indispensable for morality. In order to judge normatively an action that we ourselves practiced,

---

<sup>82</sup> Prinz, 2013.

<sup>83</sup> Darwin believed that the emergence of morality was inevitable whenever an animal had strong social instincts combined with an intellectual capacity as developed as that of humans (2009, p. 71–72).

<sup>84</sup> Joyce, 2006.

<sup>85</sup> Prinz, 2013, p. 107; Joyce, 2006, p. 50.

we must be able to identify the reasons that led us to practice it and to reflect on these reasons. That is, we must be able to compare them to other reasons that could have motivated us to act differently. Thinking in a normative way requires being aware of the grounds of our beliefs and actions, and this implies self-consciousness because it involves the capacity to identify ourselves as the subjects of our mental representations. A being without self-consciousness can be aware of the existence of an object she wants and act upon this information. On the other hand, a being with self-consciousness is also aware of the fact that *she* wants the object. She does not think just about the object she wants, but also about her own desires that make her willing to act in certain way. This self-consciousness about motives ensures a reflective distance that allows the subject to question her own motives.<sup>86</sup> Through self-consciousness, the individual is able to assume the position of a spectator of her own desires. From that position, she can compare her past and future actions and approve or disapprove of them. She may think it would have been better if she had acted according to another desire.<sup>87</sup>

This kind of thinking requires other skills, such as memory and language. The capacity to reflect on the motives that led to an action depends on the capacity to remember the action, its motives, and its results. The memory that an action provoked a feeling of dissatisfaction can, for instance, help us to form the judgement that we *ought* to have acted otherwise.<sup>88</sup> Language, on the other hand, besides allowing conscious reflection on the grounds of our actions and the formulation of the belief that we ought to have done otherwise, also allows us to share these reflections and beliefs with other members of our group. As shown in the social intuitionist model developed by Haidt, this kind of social interaction is another important element in the formation of morality.<sup>89</sup>

Throughout this paper, we have discussed the various capacities involved in moral judgements: pattern recognition, mind-reading, self-consciousness, language, memory, emotions, etc. These capacities are used in the formulation of moral judgements, but are not dedicated solely to this task. Thus, it is quite plausible that the ability to make moral judgements has arisen as a secondary effect of other mental capacities. Nevertheless, this does not mean that the capacities involved in moral judgements have not been modified as a result of their contribution to these judgements. At some moment, morality may have worked as an exaptation, i.e., as a feature that enhances fitness, but that has not evolved as a result of selective pressures related to its current role. The capacity to make moral judgements exists in different degrees, depending on the possible combinations of the elements that constitute this capacity. Certain compositions

---

<sup>86</sup> Korsgaard, 2006, p. 112–116.

<sup>87</sup> Darwin, 2009, p. 73 and 91.

<sup>88</sup> Darwin, 2009, p. 72.

<sup>89</sup> Haidt, 2012.

of these elements are more adaptive than others and can be selected. Thus, even if morality emerged as a side effect, it might have undergone subsequent adaptive structural changes, possibly because of interactions with cultural changes. Morality might be a *secondary adaptation*, as long as its elements have been modified by natural selection in consequence of the fact that the ways they interact have an impact on the fitness of individuals or groups.<sup>90</sup>

Realizing that morality involves all elements previously mentioned makes clear the difficulty of discerning whether it is an adaptation or just a *spandrel*—i.e., a by-product of the evolution of some other adaptive trait—and illustrates the shortcomings of the main approaches adopted so far.

On the one hand, we have the theory of UMG, which aims to explain some details of moral judgements, but fails to provide a plausible explanation because of the major differences between the processes involved in morality and in the language faculty. On the other hand, we have the moral foundation theory, which, in its mild version, highlights the influence of emotions on our moral judgements, but fails to provide a detailed account of many elements required for the development of morality.

In this sense, neither the evolutionary, nor the developmental theories offer a complete explanation of morality. From the developmental perspective, in order to explain what is innate in morality, we would have to detail the capacities involved, how they interact, and how they distinguish themselves from proximate cultural causes. On the other hand, being able to explain if morality is an adaptation would also involve the investigation of how and when these capabilities have emerged and which selective pressures have acted in the evolutionary process. In this case, we should still investigate to see if these pressures were sufficient to modify, by natural selection, the capacity to make moral judgements. Up to now, the authors who have studied the evolution of morality have not done much more than an investigation about the evolution of cooperation. However, a full account of the evolution of morality should also involve a story about the evolution of consciousness (or self-consciousness) as a condition for self-control. By means of self-consciousness, the agent becomes able to decide whether to adopt a purpose, and to judge it as good or bad.

---

<sup>90</sup> Joyce, 2014, p. 127–128.

## References

Ayala, Francisco J.

2010 “Colloquium Paper: The Difference of Being Human: Morality.” *Proceedings of the National Academy of Sciences of the United States of America* 107 Suppl: 9015–9022.

Blair, Richard, Robert James, Lawrence Jones, Fiona Clark, and Margaret Smith

1997 “The Psychopathic Individual: A Lack of Responsiveness to Distress Cues?” *Psychophysiology* 34 (2): 192–198.

- Blair, Richard, K.S. Pescharadt, S. Budhani, D.G. V. Mitchell, and D.S. Pine  
2006 "The Development of Psychopathy." *Journal of Child Psychology and Psychiatry* 47 (3–4): 262–276.
- Brown, Donald E.  
1991 *Human Universals*. Philadelphia: Temple University Press.
- Chomsky, Noam  
1965 *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Darwin, Charles  
2009 *The Descent of Man and Selection in Relation to Sex*. Vol. 1. New York: Cambridge University Press.
- Dupoux, Emmanuel, and Pierre Jacob  
2007 "Universal Moral Grammar: A Critical Appraisal." *Trends in Cognitive Sciences* 11 (9): 373–378.
- Dwyer, Susan  
2007 "How Good Is the Linguistic Analogy?" In *The Innate Mind: Culture and Cognition*. Vol. 2. Peter Carruthers, Stephen Laurence, and Stephen Stich, eds. Pp. 238–255. New York: Oxford University Press.
- Dwyer, Susan  
2009 "Moral Dumbfounding and the Linguistic Analogy: Methodological Implications for the Study of Moral Judgment." *Mind & Language* 24 (3): 274–296.
- Dwyer, Susan, Bryce Huebner, and Marc D. Hauser  
2010 "The Linguistic Analogy: Motivations, Results, and Speculations." *Topics in Cognitive Science* 2 (3): 486–510.
- Fessler, Daniel, Alexander Arguello, Jeannette Mekdara, and Ramon Macias  
2003 "Disgust Sensitivity and Meat Consumption: A Test of an Emotivist Account of Moral Vegetarianism." *Appetite* 41 (1): 31–41.
- Fiske, Alan Page  
1992 "The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations." *Psychological Review* 99 (4): 689–723.
- Fodor, Jerry A.  
1983 *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge: MIT Press.
- Giroux, Jessy  
2011 "The Origin of Moral Norms: A Moderate Nativist Account." *Dialogue* 50 (2): 281–306.
- Greene, Joshua, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen  
2001 "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (5537): 2105–2108.
- Haidt, Jonathan  
2001 "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814–834.

- Haidt, Jonathan  
2012 *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. 1st edition. New York: Pantheon Books.
- Haidt, Jonathan, and Fredrik Bjorklund  
2008 “Social Intuitionists Answer Six Questions About Morality.” In *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*. Vol. 2. Walter Sinnott-Armstrong, ed. Pp. 181–217. Cambridge: MIT Press.
- Haidt, Jonathan, and Craig Joseph  
2004 “Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues.” *Daedalus* 133 (4): 55–66.
- Haidt, Jonathan, and Craig Joseph  
2007 “The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules.” In *The Innate Mind: Foundation and the Future*, Vol. 3. Peter Carruthers, Stephen Laurence, and Stephen Stich, eds. Pp. 367–392. New York: Oxford University Press.
- Haidt, Jonathan, and Selin Kesebir  
2010 “Morality.” In *Handbook of Social Psychology*. 5th edition. Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey, eds. Pp. 797–832. Hoboken, NJ: Wiley.
- Hamlin, J. Kiley, Karen Wynn, and Paul Bloom  
2007 “Social Evaluation by Preverbal Infants.” *Nature* 450 (7169): 557–559.
- Harman, Gilbert  
2000 *Explaining Value and Other Essays in Moral Philosophy*. New York: Clarendon Press.
- Harman, Gilbert  
2008 “Using a Linguistic Analogy to Study Morality.” In *Moral Psychology: The Evolution of Morality: Adaptations and Innateness*. Vol. 1. Walter Sinnott-Armstrong, ed. Pp. 345–352. Cambridge: MIT Press.
- Hauser, Marc D.  
2007 *Moral Minds: The Nature of Right and Wrong*. New York: Harper Collins.
- Hauser, Marc D., Liane Young, and Fiery Cushman  
2008 “Reviving Rawls’ Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions. In *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*. Vol. 2. Walter Sinnott-Armstrong, ed. Pp. 107–143. Cambridge: MIT Press.
- Hoffman, Martin L.  
2001 *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.
- Joyce, Richard  
2006 *The Evolution of Morality*. Cambridge: MIT Press.

Joyce, Richard

- 2013 "The Many Moral Nativisms." In *Cooperation and Its Evolution*. Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser, eds. Pp. 549–572. Cambridge: MIT Press.

Joyce, Richard

- 2014 "The Origins of Moral Judgment." *Behaviour* 151 (2–3): 261–278.

Korsgaard, Christine

- 2006 "Morality and the Distinctiveness of Human Action." In *Primates and Philosophers: How Morality Evolved*. Stephen Macedo, and Josiah Ober, eds. Pp. 98–119. Princeton: Princeton University Press.

Mallon, Ron

- 2008 "Reviving Rawls's Linguistic Analogy Inside and Out." In *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*. Vol. 2. Walter Sinnott-Armstrong, ed. Pp. 145–156. Cambridge: MIT Press.

McKinnon, Margaret C., and Morris Moscovitch

- 2007 "Domain-General Contributions to Social Reasoning: Theory of Mind and Deontic Reasoning Re-Explored." *Cognition* 102 (2): 179–218.

Mikhail, John M.

- 2011 *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.

Nichols, Shaun

- 2002 "On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution." *Philosophy of Science* 69 (2): 234–255.

Nichols, Shaun

- 2005 "Innateness and Moral Psychology." In *The Innate Mind: Structure and Contents*. Vol. 1. Peter Carruthers, Stephen Laurence, and Stephen Stich, eds. Pp. 353–370. New York: Oxford University Press.

Nichols, Shaun

- 2008 "Sentimentalism Naturalized." In *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*. Vol. 2. Walter Sinnott-Armstrong, ed. Pp. 255–274. Cambridge: MIT Press.

Prinz, Jesse

- 2006 "Is the Mind Really Modular." In *Contemporary Debates in Cognitive Science*. Robert J. Stainton, ed. Pp. 22–36. Oxford: Blackwell Publishing.

Prinz, Jesse

- 2008 "Is Morality Innate." In *Moral Psychology: The Evolution of Morality: Adaptations and Innateness*. Walter Sinnott-Armstrong, ed. Pp. 367–406. Cambridge: MIT Press.

Prinz, Jesse

- 2012 *Beyond Human Nature: How Culture and Experience Shape Our Lives*. London: Allen Lane.

- Prinz, Jesse  
 2013 “Where Do Morals Come From?—A Plea for a Cultural Approach.” In *Empirically Informed Ethics: Morality between Facts and Norms*. Markus Christen, Carel van Schaik, Johannes Fischer, Markus Huppenbauer, and Carmen Tanner, eds. Pp. 99–118. Berlin: Springer.
- Roedder, Erica, and Gilbert Harman  
 2010 “Linguistics and Moral Theory.” In *The Moral Psychology Handbook*. John Michael Doris, ed. Pp. 273–296. Oxford: Oxford University Press.
- Ruse, Michael, and Edward O. Wilson  
 1986 “Moral Philosophy as Applied Science.” *Philosophy* 61 (236): 173–192.
- Schnall, Simone, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan  
 2008 “Disgust as Embodied Moral Judgment.” *Personality and Social Psychology Bulletin* 34 (8): 1096–1109.
- Schwartz, Shalom H., and Wolfgang Bilsky  
 1990 “Toward a Theory of the Universal Content and Structure of Values: Extensions and Cross-Cultural Replications.” *Journal of Personality and Social Psychology* 58 (5): 878–891.
- Shweder, Richard A., Nancy C. Much, Manamohan Mahapatra, and Lawrence Park  
 1997 “The ‘Big Three’ of Morality (Autonomy, Community, Divinity) and the ‘Big Three’ Explanations of Suffering.” In *Morality and Health*. Allan M. Brandt and Paul Rozin, eds. Pp. 119–169. New York: Routledge.
- Sperber, Dan  
 2005 “Modularity and Relevance: How Can a Massively Modular Mind Be Flexible and Context-Sensitive?” In *The Innate Mind: Structure and Contents*. Vol. 1. Peter Carruthers, Stephen Laurence, and Stephen Stich, eds. Pp. 53–68. New York: Oxford University Press.
- Sripada, Chandra  
 2008 “Nativism and Moral Psychology: Three Models of the Innate Structure That Shapes the Contents of Moral Norms.” In *Moral Psychology: The Evolution of Morality: Adaptations and Innateness*. Vol. 1. W. Sinnott-Armstrong, ed. Pp. 319–343. Cambridge: MIT Press.
- Sterelny, Kim  
 2010 “Moral Nativism: A Sceptical Response.” *Mind & Language* 25 (3): 279–297.
- Sterelny, Kim  
 2012 *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge: MIT Press.
- Suhler, Christopher L., and Patricia Churchland  
 2011 “Can Innate, Modular ‘Foundations’ Explain Morality? Challenges for Haidt’s Moral Foundations Theory.” *Journal of Cognitive Neuroscience* 23 (9): 2103–2116.

de Waal, Frans

1996 *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard University Press.

Wheatley, Thalia, and Jonathan Haidt

2005 "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16 (10): 780–784.

Whewell, William

2012 *The Philosophy of the Inductive Sciences*. Vol. 2. Seattle: Amazon Digital Services.