Chapter 8
Humean Moral Motivation

Andrés Luco

8.1 Introduction: The Problem of Moral Motivation

One of the more redeeming features of human nature is that people are capable of doing the right thing – of acting with exceptional benevolence and justice. Philosophers have long been trying to understand the motivations that underlie moral conduct, and for good reason. Discoveries about the facts of moral motivation could overturn some of the most fundamental postulates of moral philosophy. Suppose *psychological egoism* turns out to be true. Psychological egoists hold that all human behaviour is motivated ultimately by self-interest. But as Russ Shafer-Landau has noted, if psychological egoism is true, then altruism is impossible (Shafer-Landau 2010, pp. 88–89). And yet, many of the moral duties we have seem to require altruism. Surely, however, we cannot have moral duties to do the impossible. Thus, if psychological egoism is true, it may entail the unpalatable consequence that there are no moral duties which require us to act altruistically. Another rationale for philosophical interest in moral motivation has to do with the practical aims of ethics. It is all very fine to theorize about the right and the good, but we would of course like to see people actually *do* right and *pursue* the good. Accordingly, ethicists from Plato to Rawls have put considerable effort into understanding the levers of human psychology which can be pulled to generate ethical behaviour.

In this essay, I shall defend the *Humean theory of motivation* (hereafter 'Humeanism') as the best account of moral motivation. Humeanism is often called the 'belief–desire model' of action because it explains all intentional behaviour – including moral behaviour – by citing a combination of a desire for some end and a belief that an action is a means to achieve that desired end. I offer three arguments in favour of Humeanism. First, I suggest that Humeanism meets one criterion

---

A. Luco (✉)
Philosophy Group, Nanyang Technological University, Singapore, Singapore
e-mail: acluco@gmail.com

B. Musschenga and A. van Harskamp (eds.), *What Makes Us Moral? On the capacities and conditions for being moral*, Library of Ethics and Applied Philosophy 31,
DOI 10.1007/978-94-007-6343-2_8,

of theory choice better than its theoretical competitors. This is the *continuity constraint*. Humeanism is consistent with a *continuous natural history of moral motivation*, i.e. an evolutionary account which explains how the moral motivations of modern humans evolved from simpler precursors resembling the psychological traits of nonhuman animals (especially nonhuman primates). The same cannot be said for the rivals of Humeanism. Then, in order to ward off the objection that nonhuman primates are too different from humans to provide any insight into moral motivation, I offer a second argument for Humeanism. It draws on another criterion of theory choice called *Morgan's Canon* – a standard of theoretical parsimony according to which a psychological mechanism should be attributed to an organism only if it's the *sole* mechanism which can cause some behaviour. I contend that Humeanism meets this criterion more successfully than rival accounts of moral motivation. I will so argue on the grounds that the Humean belief–desire model can successfully explain both moral action and a variety of non-moral actions. Finally, I shall argue that recent findings in neuroeconomics furnish empirical evidence in favour of Humeanism, and against anti-Humeanism.

8.2 Theories of Moral Motivation: Humeanism Versus Anti-Humeanism

In this section I describe the dominant philosophical theories of moral motivation. First, I canvass a few preliminary observations and assumptions about moral motivation in general. *Moral action* is action *judged by the agent* to be either morally praiseworthy or required. By this definition, an agent's own moral judgments determine whether his or his actions count as *moral* actions. The definition does not presuppose that an agent must *correctly* judge his or her actions to be morally praiseworthy or required in order for those actions to be moral actions. *Moral motivation* refers to the motivations that explain moral action. As Connie Rosati notes, the 'basic phenomenon' of moral motivation is that moral motivations reliably change in accordance with changes in moral judgment. For instance: if someone judges that φ is morally right, then he or she will ordinarily be motivated to φ; but if this person becomes convinced that φ is morally wrong and that ψ is morally right, he or she will ordinarily become motivated to ψ rather than φ (Rosati 2006, §1). It is also thought that moral motivation varies with moral judgment due to practical reasoning. *Practical reasoning* refers to the process of thinking about how one should act. All the theories of moral motivation canvassed below assume that moral conduct is the result of a process of practical reasoning initiated by moral judgment.

It is also commonly assumed that moral action is a species of intentional action. *Intentional action* is action motivated by the intentional states of the agent. Intentional states are psychological states that are *about*, *of*, or *for* something else. They include beliefs, desires, and emotions – e.g. a belief *about* the existence of God, a desire *for* pistachio ice cream, a fear *of* snakes, and so on. The 'content' of an intentional state is whatever the state is about, of, or for. I have a belief about the existence of God; so, the content of that belief includes the concept of God.

It will also become clear that the rival theories of moral motivation all distinguish between cognitive and affective states of mind. *Cognition* is associated with thinking, believing, reasoning, learning, decision making, and memory. *Affect*, or emotion, is linked to feeling and wanting. However, there is no consensus definition of either cognition or affect. Despite this unfortunate fact, I will rely on a promising account of the cognition–affect distinction articulated by Jesse Prinz (cf. Prinz 2004, pp. 41–51). According to Prinz,

> cognitive states and processes are those that exploit representations that are under the control of an organism rather than under the control of the environment. A representation is under organismic control if the organism has activated it or maintains it in working memory. A cognitive state is one that contains such a representation, and a cognitive process is one that activates, maintains, or manipulates such a representation. (Prinz 2004, pp. 45–46)

Prinz uses the term 'organismic control' as a synonym for 'executive control', He identifies the 'executive' brain structures as those centred in the prefrontal cortex (Prinz 2004, p. 47). The sorts of representations that Prinz takes to be under organismic control are *concepts*, as opposed to *percepts*. When you see a shape, you have a percept of that shape. But you can also store a copy of that percept in your memory. The copy of that percept recalled from memory is a concept of the shape. Thus, a cognitive state, in Prinz's view, is a mental state which contains concepts (Prinz 2004, p. 46).

Prinz's definition of a cognitive state accounts for many intuitions about what mental activities count as 'cognitive' (Prinz 2004, p. 48). For instance, cognition is associated with the conscious processing of information ('access consciousness'). Such processing includes acts of thinking, like doing long division, and deliberation about actions. It therefore involves organismic control. Also, cognition is thought to be effortful. Choosing a goal, for instance, seems to require some kind of mental work. The frontal cortex contributes to making strategic choices among competing goals. By contrast, affect occurs unbidden. You don't have to make an effort to feel sad or angry; these emotions are induced by events in the surrounding environment.

Bearing in mind the differences between cognitive and affective states, we shall see that competing accounts of moral motivation can be distinguished by the way they respond to two questions:

1. *Can a cognitive (i.e. belief-like) psychological state produce a motivation at time t, without being accompanied by an affective (i.e. emotional or desire-like) psychological state which exists prior to t?*
2. *Are moral judgments cognitive states or affective states?*

The first theory of moral motivation on the table is the Humean theory of motivation – also designated here as 'Humeanism'. The theory bears this name because its adherents claim an intellectual debt to David Hume. Humeanism is also called the *belief–desire model* of action, because it holds that intentional action is explained by a combination of a desire for some end, together with a belief that some action is a means to satisfying the desire. A belief that some action is a means to achieving some desired end is called a *means–ends belief.*

The Humean answer to the first question above is 'no' – it is not psychologically possible for a cognitive state to produce a motivation in an agent, unless it is accompanied by some existing affective state. Humeans treat beliefs as paradigmatic cognitive states, and desires as paradigmatic affective states. Additionally, Humeans assume that cognitive and affective states are mutually exclusive. Thus, Humeans hold that no belief is a desire. The basic Humean position with respect to the first question, then, is that no *belief* can motivate intentional action without being accompanied by an existing desire.

Regarding the second question, Humeans are split into two camps: cognitivists and non-cognitivists. *Cognitivism* is the view that a moral judgment expresses a cognitive state – in particular, a belief. The defining feature of beliefs is that they are *truth-apt*, meaning they are capable of being either true or false. *Non-cognitivism*, by contrast, is the view that a moral judgment expresses a non-cognitive state, such as an emotion or desire. Non-cognitive states are supposedly not truth-apt. Desires, for instance, cannot be true or false – they can only be satisfied or unsatisfied. Non-cognitivist Humeans include such authors as Simon Blackburn (1984) and Allan Gibbard (1990), whereas proponents of cognitivist Humeanism include John Mackie (1977) and Nick Zangwill (2003).

Humeans maintain that desires play a role quite independently of beliefs in generating moral motivation (Zangwill 2008a, b). At least three kinds of desires can contribute to moral motivation: (1) self-regarding desires, (2) other-regarding desires, and (3) moral desires. First, agent *S* has a *self-regarding desire* to achieve some end *E*, if *E* is exclusively a state of *S*. Second, agent *S* has an *other-regarding desire* to achieve *E* if *E* is exclusively a state of some person or group *other than S*. For instance, suppose that I donate blood purely out of a desire to win the respect of my peers through conspicuous acts of benevolence. If you were to ask me, I would honestly admit that if donating blood did not do wonders for my reputation, I wouldn't have any desire to do it. In this case, I have a self-regarding desire to donate blood, where the desired end is an improved reputation. An improved reputation is a state in which I find myself; it is not a state of anyone else. On the other hand, suppose instead that I donate blood because of a desire to promote the health of people I've never met. Suppose my desired end is to change the states that other people are in, and not my own state, by improving the health of those in need of blood transfusions. To the extent that my end is exclusively to improve the health of others, I have an other-regarding desire to donate blood.

Finally, and perhaps most importantly, moral motivation can result from moral desires. Agent *S* has a *moral desire* to achieve end *E* if *S* judges that *E* bears a set of moral properties, and *S* desires the moral properties of *E*. For instance, I will have a moral desire to donate blood if I judge that it is morally right to donate blood, and I want to do what I judge to be right. The content of a moral desire can be understood as a concept or representation of moral properties such as *rightness*, *goodness*, *praiseworthiness*, etc. Conversely, the content of a *non-moral* desire would not include any representation of moral properties. In this essay, no position is taken on the existence of mind-independent moral properties. Even a moral error theorist could assent to the definition of moral desires just introduced; but would simply add the caveat that the contents of moral desires are *moral concepts* that only *purport* to represent moral properties. However, the error theorist will add that moral concepts *misrepresent* reality, because moral properties do not exist. Likewise, a non-cognitivist Humean like Blackburn could accept the foregoing definition of moral desires. But he would specify that moral properties are projections of sentiment, and that no mind-independent moral properties exist.

The second group of theories about moral motivation is classified under the heading of *anti-Humeanism*. With regard to the first question above, the anti-Humean answer is 'yes'. Anti-Humeans claim that a cognitive state, or a state with both cognitive and affective features, can be sufficient to generate motivation without being paired up with any independently existing affective state. Regarding the second question, anti-Humeans treat moral judgments as either cognitive psychological states, or as hybrid psychological states with both cognitive and affective features.

Margaret Olivia Little (1997) defends two prominent versions of anti-Humeanism. Each version advances a different theory of the cognitive (or semi-cognitive) psychological states that motivate moral action. According to the first version of anti-Humeanism, moral judgments are partly cognitive, and partly non-cognitive; they are hybrid psychological states that have both cognitive (belief-like) and affective (desire-like) features. Accordingly, these hybrid states have been dubbed '*besires*' (Little 1997, p. 254).[1] Besires are claimed to be both belief-like and desire-like because they supposedly have both of two *directions of fit* that distinguish beliefs from desires. Beliefs have a 'mind-to-world' direction of fit, because they are the sort of intentional state that goes out of existence when the subject perceives evidence that the world is not how he believes it to be. If, for instance, I believe that it's 9 am on Monday, but my friend informs me that it's actually 10 am, my belief that it's 9 am will be extinguished because it is contradicted by my friend's (presumably reliable) testimony. On the other hand, desires have a 'world-to-mind' direction of fit, since the content of a desire is a representation of how the world *is to be*. Whenever there is a discrepancy between how the world is and how the subject wants the world to be, a desire has the effect of motivating the subject to *change* the world until his desire is satisfied (cf. Smith 1987, p. 56).

Besires are said to have both directions of fit. David McNaughton, another anti-Humean, suggests that moral judgments are besires. A moral judgment has a mind-to-world direction of fit, in the sense that an agent can perceive the moral properties of actions, persons, institutions, etc. For instance, one can perceive that an action is morally required (McNaughton 1988, p. 109). But at the same time, a moral judgment has a world-to-mind direction of fit, i.e. a moral judgment can motivate an agent to change the world and carry out what he judges to be morally required.

---

[1] The term 'besire' was coined by J.E.J. Altham (1986).

Another group of anti-Humeans postulate a different kind of cognitive state to explain moral motivation. Little calls them *desire-entailing beliefs* (Little 1997, pp. 254, 259–260). According to the anti-Humeans in question, moral judgments are beliefs, and at least some moral beliefs are desire-entailing beliefs (Little 1997, p. 261). Moral beliefs 'entail' desires in the sense that 'possession of certain desires is a necessary condition for possession of certain moral beliefs' (Little 1997, p. 260). Here it's crucial to bear in mind that the explanatory direction of the entailment is *from* moral beliefs *to* desires. Possessing certain moral beliefs *necessarily explains* the formation of certain desires to act in accordance with those beliefs. Thus, if one lacks the desires which would be entailed by a desire-entailing moral belief, either one does not genuinely have the relevant moral belief in the first place, or one is afflicted by some form of irrationality which interferes with the normal process of entailment (Little 1997, pp. 260–261; Zangwill 2008b, pp. 96–97, 111). For instance: suppose that I believe I have a moral duty to minimize avoidable suffering, and I notice that a shortage of blood supplies causes avoidable suffering. If my moral belief is a desire-entailing belief, I must form the desire to donate blood *because* of that moral belief.

Unlike besires, desire-entailing beliefs are not hybrid states with dual directions of fit. They are ordinary beliefs with mind-to-world direction of fit. Also, the desires entailed by desire-entailing beliefs are ordinary desires with world-to-mind direction of fit. However, the idea of a desire-entailing belief is contrary to Humeanism. It suggests that a belief unaccompanied by an existing desire state is by itself sufficient to produce a new desire. Humeans think this is impossible.

Now I turn to three arguments in favour of the Humean account of moral motivation. My remarks are not designed to favour either cognitivist or non-cognitivist Humeanism. Instead, the arguments are intended to provide a general case for Humeanism and against anti-Humeanism.

8.3 For Humeanism: The Argument from Continuity

Humeanism and anti-Humeanism are rival theories in the psychology of motivation. As such, they are subject to the same criteria of adequacy as any psychological hypothesis. In this section I suggest that Humeanism meets one theoretical criterion better than anti-Humeanism. This criterion derives from the theory of evolution, and I will call it the *continuity constraint*:

Continuity Constraint

> Given that human beings and other animals evolved from common ancestors, a theory of human psychology should be consistent with the best available explanation of how human psychological traits evolved from simpler precursors resembling the psychological traits of nonhuman animals (especially nonhuman primates).

The continuity constraint is based on one of the most profound insights of evolutionary theory. It presupposes that organic structures, including psychological structures, are products of a slow, incremental process of ‘descent with modification’, in which newer, more complex structures arise as piecemeal modifications of older, often simpler ones.[2] The continuity constraint advises us to prefer psychological theories which are consistent with explanations of how human psychological traits evolved from simpler antecedents resembling the psychological traits of nonhuman animals. The psychologies of nonhumans – especially nonhuman *primates* – furnish important sources of evidence for the motivations that were probably characteristic of (now extinct) ancestor species we modern humans have in common with the other animals.

Many significant philosophers and scientists have defended the continuity constraint. Darwin himself argued for the psychological continuity of humans and nonhumans (cf. Sober 1998, p. 228). In the second edition of *The Descent of Man* (1874), Darwin wrote:

> The difference in mind between man and the higher animals, great as it is, certainly is one of degree and not of kind ... the senses and intuitions, the various emotions and faculties, such as love, memory, attention, curiosity, imitation, reason, &c., of which man boasts, may be found in an incipient, or even sometimes in a well-developed condition, in the lower animals. (Thompson 1995, p. 71)

In the same treatise, Darwin suggests that the precursors of human moral psychology could be found in other social animals:

> Any animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well, or nearly as well developed, as in man. (Thompson 1995, pp. 41–42)

Daniel Dennett also defends a commitment to the continuity constraint. He urges that the continuity constraint applies even when an organism exhibits unique and unprecedented traits, as is the case with modern humans. Even when the emergence of a new organism or biological structure at some moment in evolutionary history appears radically different from what existed previously, Dennett maintains that theorists should search for

> a subprocess or special feature of a design process that can be demonstrated to permit the local speeding up of the basic, slow process of natural selection, *and* that can be demonstrated to be itself the predictable (or retrospectively explicable) product of the basic process. (Dennett 1995, p. 76)

In sum, the continuity constraint requires that the origins of an observed psychological trait should be explicable as a product of descent with modification from

---

[2] Evolutionary change does not *only* proceed from the simple to the complex. As Elliot Sober notes, ‘the history of life is peppered with cases of evolutionary simplification. For example, the evolution of parasites typically involves a transition from complex to simple, as the parasite loses features of its free-living ancestor’ (Sober 1998, p. 225). When it comes to understanding the evolutionary origins of human psychological features, though, we obviously seek an explanation for how something as complex as the human mind arose from simpler precursors.

pre-existing traits. Since the theory of evolution implies that humans and other animals share common ancestors, the older traits on which natural selection operated to produce the psychological capacities of modern humans can probably be found in nonhuman creatures – most especially, the extant nonhuman primates. Even when a trait appears to be radically new and different, it should still be explicable as the product of evolutionary processes operating on pre-existing traits.

I argue that Humeanism meets the continuity constraint, while anti-Humeanism does not. The motivations which Humeans posit to account for moral action are the same as, or very similar to, the motivations that researchers have invoked to explain what I call *proto-moral behaviours* in nonhuman primates. Human beings are capable of *altruism*, i.e. of helping and caring for others. We extend aid to our family members, friends, neighbours, and even strangers, sometimes at great sacrifice to ourselves. When we ask ourselves why we help and care for others the way we do, we often resort to moral language – we speak of the *duties* of a good mother or son, the *virtues* of a loyal friend, and so on. Thus, some altruistic behaviours, at least when carried out by human beings, are instances of *moral* action. But some nonhuman animals – particularly the other primates – engage in similar forms of helping and caring. Such animal behaviour can be called *proto-moral*, since it may furnish insight into the evolutionary precursors of an important class of morally significant actions exhibited by humans. The continuity constraint supports the expectation that the origins of the psychological capacities associated with human moral motivation should be traceable back to ancestors that *homo sapiens* shares with other primates. As we shall see, Humeanism meets the continuity constraint because it explains how proto-moral motivations could have been shaped by natural selection into the fully fledged moral motivations of modern humans. The same cannot be said for anti-Humeanism.

Primatologist Frans de Waal has observed compelling parallels between helping/caring behaviours in human beings and nonhuman primates. In fact, De Waal himself voices a commitment to the continuity constraint. He argues that some primates – particularly the great apes – are motivated by the same psychological mechanisms which underlie human propensities for empathy and altruism (De Waal 2006, p. 28, 2008). In particular, De Waal asserts the following:

> In human behavior, there exists a tight relation between empathy and sympathy, and their expression in psychological altruism . . . It is reasonable to assume that the altruistic and caring responses of other animals, especially mammals, rest on similar mechanisms. (De Waal 2006, p. 28)

De Waal characterizes empathy and altruism as necessary conditions for moral conduct. Furthermore, he holds that empathy and altruism are fundamentally affective motivations based on an ‘emotional interest in others’ (De Waal 2006, pp. 20–21). In line with the continuity constraint, De Waal contends that the same affective states which generate moral motivation in humans also motivate proto-moral behaviour in nonhumans.

De Waal uses the term ‘*intentional altruism*’ to denote the suite of motivations which explain helping behaviour (De Waal 2008, p. 281). Intentional altruism consists of *intentional states* that motivate helping. De Waal argues that *empathy* is the proximate mechanism through which intentional altruism operates (cf. De Waal 2008, p. 282). Empathy is defined by De Waal as 'the capacity to (*a*) be affected by and share the emotional state of another, (*b*) assess the reasons for the other's state, and (*c*) identify with the other, adopting his or her perspective' (De Waal 2008, p. 281).

De Waal proposes what he calls a 'Russian Doll Model' of empathy. The model consists of three levels of empathy: (1) *emotional contagion*, (2) *sympathetic concern*, and (3) *empathetic perspective-taking* (De Waal 2008, pp. 287–288). These three levels of empathy are distinguished according to the complexity of the psychological mechanisms needed for an organism to experience each level. Emotional contagion – the lowest level of empathy – can be attributed to rats, birds, and monkeys. These animals are much less psychologically complex than large-brained mammals like cetaceans and the great apes. Emotional contagion consists of the matching of emotional states between one individual (the 'subject') and another (the 'object') (De Waal 2008, p. 282). For example, mice exhibit intensified pain responses when they see other mice exhibiting similar responses. Also, rhesus macaque monkeys tend to terminate projected pictures of conspecifics in a fearful pose (De Waal 2008, pp. 283, 288). In each of these examples, the emotional state of one individual 'spreads' to another.

The second level of empathy is *sympathetic concern*, which involves emotional concern about a distressed or needy other (De Waal 2008, p. 283). Additionally, the sympathizing subject's concern motivates attempts to ameliorate the other's distress. A manifestation of sympathetic concern is consolation behaviour, which de Waal defines as 'reassurance provided by an uninvolved bystander to one of the combatants in a preceding aggressive incident' (De Waal 2006, p. 33). For example, De Waal observed a juvenile chimpanzee approach and put its arms around a screaming adult male who had just been defeated in a fight. Quantitative studies have found that in chimpanzees, third parties direct consolation behaviour more at recipients of aggression than at aggressors, and more at recipients of *intense* aggression than mild aggression (De Waal 2006, p. 35).

De Waal takes sympathetic concern to involve more complex psychological capacities than emotional contagion, for it involves both an emotional concern which is sensitive to the emotional state of another being, and a motivational disposition to try and reduce the distress of the other. Animals capable of emotional contagion, but not sympathetic concern, do not engage in consolation behaviour. For instance, macaque monkeys do not even comfort their own offspring after a fight. By contrast, the reassurance of distressed others is typical in the great apes, which are closer phylogenetic relatives to modern humans than monkeys (De Waal 2008, p. 285).

De Waal's third level of empathy is *empathetic perspective-taking* (henceforth EPT). According to De Waal, EPT consists in 'the capacity to take another's perspective, e.g. understanding another's specific situation and needs separate from one's own, combined with vicarious emotional arousal' (De Waal 2008, p. 285). Like the first two levels of empathy, EPT involves emotional engagement with the other, along with a disposition to reduce the other's distress. However, EPT is more complex because it requires the ability to attribute mental states to the other, and to recognize the causes of those states. The best evidence for EPT is the phenomenon of *targeted helping*, which is helping behaviour fine-tuned to the needs and situation of the other (De Waal 2008, p. 285). De Waal cites hundreds of observations of targeted helping in apes, dolphins, and elephants. For instance, researchers have frequently observed that when juvenile orangutans get stuck in a tree, they are rescued by their mothers. The mother orangutans drape their bodies between one branch and another, thereby creating a bridge to safety for their whimpering offspring. This behaviour clearly requires an understanding of the causes of the juvenile's emotional distress. De Waal holds that the mechanisms and cognitive abilities associated with empathetic perspective-taking must be complex. For this reason, EPT is exhibited by only a few large-brained animal species other than humans – particularly the great apes, elephants, and dolphins (De Waal 2008, p. 286).

De Waal argues that the more complex levels of empathy are descended from, and even enabled by, the simpler levels (De Waal 2008, pp. 287–288). He hypothesizes that EPT evolved through processes of reciprocity and kin selection from simpler mechanisms related to emotional contagion. Emotional contagion and sympathetic concern could enable the organism to share in the emotional state of the distressed subject while developing a motivation to alleviate that distress. The work left for EPT, then, would be to identify the cause of the subject's distress and to determine how to alleviate it. Hence, De Waal suggests that the capacity for EPT would have been a relatively new trait that emerged with the common ancestor of humans and apes, whereas the capacity for emotional contagion was inherent in our common ancestor with monkeys and probably in far more ancient organisms (De Waal 2008, p. 292).

The mechanisms that De Waal theorizes to explain directed altruism in nonhuman primates, and ultimately in humans as well, are consistent with the Humean theory of motivation. For instance, a chimpanzee with the capacity for empathetic perspective-taking shares in the emotional state of the one in need or distress, assesses the causes of the need or distress, and then takes the means necessary to meet the needs of the other. The chimp can be said to be motivated by Humean motivational mechanisms, including a desire to meet the needs of a conspecific, and a belief about the means necessary to meet those needs.

In contrast, De Waal's Russian Doll Model is not consistent with the motivations postulated by anti-Humeanism. Anti-Humean mechanisms depend on a level of cognitive sophistication that is far beyond that of any nonhuman animal. As we saw, both desire-entailing beliefs and besires require an ability to make moral judgments, i.e. to apply moral concepts such as *forbidden*, *permissible*, *obligatory*, *virtuous*, *vicious*, *just*, and *unjust* to actions, persons, and states of affairs. But nonhuman animals do not have moral concepts. Richard Joyce convincingly argues that nonhumans lack moral concepts, because they cannot carry out a 'semantic ascent' through which they regard certain categories of actions as *worthy* or *unworthy* of acceptance, rather than merely *accepted* or *not accepted* (Joyce 2006, pp. 82–85). So, the cognitive limitations of nonhuman animals render them incapable of making moral judgments. If nonhumans cannot form judgments about what is morally required of them, or about what behaviour is worthy or unworthy of acceptance, they could not be motivated to act on such judgments. Thus, besires and desire-entailing beliefs could not have a place in the psychologies of even the most intelligent nonhumans.

On the other hand, it may be tempting to interpret some of De Waal's experimental findings as evidence that primates have primitive moral concepts. For instance, when a capuchin monkey is given a cucumber and a neighbouring capuchin is given a more attractive grape, the first monkey often reacts negatively, e.g. by throwing the cucumber (cf. De Waal 2006, pp. 44–49). However, Sarah F. Brosnan suggests that there is insufficient evidence to justify attributing such reactions to a 'sense of fairness'. She points out that it is impossible to tell whether nonhuman animals have an understanding of fairness, because there is no way to get them to describe the nature of that understanding (Brosnan 2010, p. 80).

If anti-Humean motivations exist, they would have appeared after the evolution of moral judgment in either modern human beings or in some hominid ancestor of *homo sapiens*. Given Joyce's argument that a semantic ascent is necessary for possession of moral concepts, a hominid ancestor species capable of moral judgment would need a robust capacity for language. Thus, anti-Humeanism implies a discontinuity in evolutionary history. It suggests that a new trait – either besires or desire-entailing beliefs – would have emerged without identifiable precursors. The continuity constraint counsels us to minimize such discontinuities, provided that it does not result in explanatory loss. Even where evolutionary discontinuities are necessary to explain what is observed, the continuity constraint teaches us to show that they are products of some evolutionary process. It is far from clear why desire-entailing beliefs or besires would have been selected for. However, perhaps the anti-Humean motivations did not themselves evolve from any process of selection. They may have been byproducts ('spandrels') of other selected traits that are unique to modern humans, such as large brain size or cerebral complexity. But even if that's the case, it is not clear how besires or desire-entailing beliefs could have been a *predictable*, or *retrospectively explicable*, byproduct of some selective force operating on pre-existing traits. Spandrels may not be selected for, but they can still be explained as products of selection for other traits. Ultimately, beliefs and desires, as Humeans understand them, stand a far better chance of being included in a continuous evolutionary etiology of moral motivation. For this reason, the continuity constraint supports a presumption in favour of Humeanism.

8.4 For Humeanism: The Argument from Morgan's Canon

For all that has been said about continuity, human beings are probably unique among the animals in possessing the ability to make moral judgments. An anti-Humean may point out that the argument from continuity overemphasizes the extent to which humans are *descended* from ape-like ancestors, while it underplays the extent to which we are marvellously complex *modifications* of our ancestors. In answering this objection, I shall appeal to another criterion for theory choice to show that there is no good reason to suppose that moral action must be explained by anti-Humean mechanisms which humans do not share with other animals. This principle is known as *Morgan's Canon*. It is so named after comparative psychologist C. Lloyd Morgan. In his *Introduction to Comparative Psychology* (1894), Morgan states his principle as follows:

> In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercises of one which stands lower in the psychological scale. (quoted in Sober 1998, p. 224)

Morgan's Canon applies to theories that explain a creature's behaviour by citing its internal psychological mechanisms. This is, of course, how the dominant theories of moral motivation purport to explain moral action. It's quite appropriate, therefore, to evaluate theories of moral motivation with respect to Morgan's Canon.

Although Morgan himself believed his principle to be justified by Darwinian considerations, Elliot Sober argues that it can be substantiated by deductive logic. To demonstrate this, Sober offers a novel interpretation of Morgan's Canon, summarized as follows (Sober 1998, pp. 236–237):

> **Morgan's Canon**
>
> Let one internal mechanism, H, be higher than another, L, if and only if the behavioral capacities entailed by H properly include the behavioral capacities entailed by L. Suppose, for instance, that L entails the behaviors in set B1 and H entails the behaviors in both sets B1 and B2 (where B1 and B2 are not equivalent sets).
>
> Then:
>
> if an organism performs the behaviors in B1, but not the behaviors in B2, the organism lacks H.

On Sober's interpretation of Morgan's Canon, so-called 'higher' psychological mechanisms enable an organism to do more than 'lower' ones. Accordingly, *H* is a higher mechanism than *L*, because an organism with *H* can perform the behaviours in both sets *B1* and *B2*, whereas an organism with only *L* can only perform the behaviours in set *B1*.

So construed, Morgan's Canon states that attributions of psychological mechanisms are constrained by the following deductive inference:

1. Mechanism *H* entails the behaviours in *B1* and *B2*.
2. Organism *O* does not display the behaviours in *B2*.
3. Therefore, by *modus tollens*, *O* does not have *H*.

Thus, Morgan's Canon advises us to attribute a psychological mechanism to an organism only if it's the *sole* mechanism which can explain some behaviour. This is not to say that all behaviours of the same type must be explained by *the same* mechanism. To use Sober's example, stickleback fish and chimpanzees both build nests, but Morgan's Canon would not imply that the same psychological process must motivate nest-building in both species (Sober 1998, p. 230). Instead, the

Canon allows the psychological process which motivates nest-building in chimps to be different, but only if this process *alone* could explain additional behaviours exhibited by chimps and not by sticklebacks.

Sober offers an example to illustrate how Morgan's Canon should guide theory choice in comparative psychology. When a species of bird called the piping plover sees a predator approach its nest, it moves away from the nest, sits down, and starts peeping loudly. In so doing, the bird makes it appear as though it's sitting on eggs, when really it isn't (Sober 1998, p. 238). This behaviour has been called 'false nesting'. What explains this curious behaviour? Consider two possibilities. Let *H1* be the hypothesis that piping plovers want to protect their young, and they believe false nesting will accomplish this without knowing exactly *how*. Let *H2* be the hypothesis that the plovers want to protect their young, and they believe false nesting will accomplish this *by* causing the predator to believe that the eggs are not in their true location. Unlike *H1*, *H2* attributes *second-order intentionality* to the plovers. Second-order intentionality is the capacity to form intentional states (e.g. beliefs or desires) *about* the intentional states of others. If piping plovers have second-order intentionality, then a plover can form the belief that the false nesting display will make an approaching predator incorrectly believe that the plover is sitting on its eggs. However, Morgan's Canon instructs us to reject *H2* in favour of *H1*, since the plover does not behave in ways that are explained *only* by second-order intentionality. False nesting can be exhaustively explained by *H1*, the hypothesis of *first-order intentionality*. The plover may simply believe that the false nesting display will lure the predator away from the true location of the eggs, without believing that the display will lure away the predator *by* manipulating the predator's beliefs about the location of the eggs. By contrast, the plover does not exhibit any behaviours which are *uniquely* explained by second-order intentionality. And so, by Morgan's Canon, it should be concluded that the piping plover lacks second-order intentionality.

Morgan's Canon can be deployed in an argument against anti-Humeanism. The argument begins with an observation that is hard to dispute: the Humean belief–desire model can successfully explain at least *some* human behaviours. When explaining the actions of other people, we often appeal to the beliefs and desires that motivated them. Philosophers call this explanatory strategy *folk psychology*. Folk psychology is commonly deployed in explanations of *non-moral* action. By 'non-moral action' I mean action which is not morally significant. Consider the act of tying your shoe. You are under no *moral* obligation to tie your shoe. Neither is it morally wrong, nor morally praiseworthy, nor blameworthy for you to tie your shoe. In and of itself, shoe-tying is a non-moral act. Folk psychology provides a straightforward explanation for your decision to tie your shoe: you have a desire to walk about with shoes on, and you believe that tying your shoe will enable you to do so successfully. According to Morgan's Canon, a 'higher' psychological mechanism should be attributed to an agent only if it can explain behaviours that a 'lower' mechanism could not. If we follow the Canon, then the burden is on the anti-Humean to show what behaviours, if any, can *only* be explained by anti-Humean mechanisms such as besires or desire-entailing beliefs. If there are no such behaviours, then Morgan's Canon would call for dispensing with the anti-Humean mechanisms.

The most likely anti-Humean rejoinder is to argue that there is no adequate way of explaining *moral* action other than by appeal to besires or desire-entailing beliefs. Yet I submit that motivational Humeanism can fully account for moral action. From a Humean perspective, moral action is explained by either a self-regarding, other-regarding, or moral desire plus a belief about how to satisfy the desire. Of these three types of desire, moral desires are the least well-understood. Accordingly, in the remainder of this section I shall elucidate the role that moral desires can play in a Humean explanation of moral action.

In my view, moral desires are induced by what Jonathan Haidt and other researchers call *moral emotions* (cf. Haidt 2003; Moll et al. 2008). According to Haidt, moral emotions are 'those emotions that are linked to the interests or welfare of society as a whole or at least of persons other than the judge or agent' (Haidt 2003). In his review of moral emotions research, Haidt says that moral emotions are generally elicited by situations and events that affect others (Haidt 2003). But judging from the details of Haidt's presentation, another critical factor which sets moral emotions apart from other emotions is their link to moral judgment. For instance, *self-critical* moral emotions, such as guilt, shame, and embarrassment, are elicited when one judges oneself to have committed some kind of moral violation, while *other-critical* moral emotions, such as indignation, anger, contempt, and disgust, are elicited when one judges someone else to have acted immorally (Haidt 2003; Moll et al. 2008, pp. 2–3, 6). Furthermore, moral emotions have effects on motivation. They are said to generate *action-tendencies,* i.e. they put a person in a motivational state in which 'there is an increased tendency to engage in certain goal-related actions (e.g., revenge, affiliation, comforting, etc.)' (Haidt 2003).

The action-tendencies of moral emotions provide evidence for the link between moral emotions and moral desires. Moral emotions have motivational effects by virtue of inducing moral desires in the agent. Consider anger. The action tendencies associated with anger include the motivation to redress situations the subject judges to be a moral violation. Evidence that anger motivates people to redress perceived moral wrongs can be drawn from an experimental economics paradigm involving the *ultimatum game*. In the ultimatum game, two players are shown a sum of money, say $10. The first player, called the 'proposer', is instructed to offer any whole number of dollars, from $1 to $10, to the second player, who is called the 'responder'. The proposer can make only one offer, and the responder can either accept or reject this offer. If the responder accepts the offer, the money is shared according to the terms of the offer. If the responder rejects the offer, neither player receives anything.

When the ultimatum game is played by people in non-repeated experimental trials and with varying amounts of money (including large sums of money), proposers most often offer 50 % of the original sum and respondents reject offers below 20 % about half of the time (Gintis et al. 2007, p. 608; Sanfey et al. 2003, p. 1755). Participants in these experiments reported that they reject low offers because they are angered by offers they judge to be unfair (Sanfey et al. 2003,

p. 1756; cf. Pillutla and Murnighan 1996). In addition, Sanfey et al. (2003) used fMRI scans on subjects playing the ultimatum game, and found that brain areas associated with anger were activated in responders by low or ‘unfair’ offers of $2:$8 or less. Activation of the bilateral anterior insula is involved in the neural realization of specific other-critical moral emotions, including disgust and anger (Sanfey et al. 2003, p. 1757; Moll et al. 2008, p. 15). Greater activations of the bilateral anterior insula were correlated positively with the degree of an offer’s unfairness. The anterior insula exhibited greater activations in response to offers of $1:$9 than to $2:$8, and offers of $2:$8 generated greater activations than ‘fair’ offers of $5:$5. Moreover, the magnitude of activation in this brain region correlated with subsequent decisions to reject the offer. Participants with stronger anterior insula activations in response to unfair offers were more likely to reject an unfair offer. Also, the anterior insula exhibited more intense activation in response to unfair offers from a human proposer than unfair offers from a computer. From these results, Sanfey and colleagues conclude that activation of the anterior insula is involved in the neural realization of anger directed at persons who intentionally violate a principle of fairness. The experience of anger in turn motivates resistance to unfairness in the form of a rejected offer (Sanfey et al. 2003, p. 1756).

The Humean can offer an intuitive explanation for Sanfey et al.’s findings: anger generated by an offer judged to be unfair induces in subjects a moral desire to reject the offer. People in Sanfey’s experiments understood themselves to be rejecting a low offer because they judged it to be unfair. So here we have instances of a moral judgment generating a moral action. Interestingly, fMRI imaging suggests that the mechanism through which judgments of unfairness motivated rejections of offers is entirely consistent with the Humean belief–desire model. Thus, the moral actions observed in Sanfey’s study are not of a sort that *only* the anti-Humean machinery of besires or desire-entailing beliefs would be able to produce. And since the anti-Humean motivations are not uniquely necessary to moral action, Morgan’s Canon would have us discard an anti-Humean account in favour of a Humean belief–desire model which is sufficient to explain both moral and non-moral behaviour.

8.5 Explanatory Deficits of Anti-Humeanism

I have been arguing that moral action can be exhaustively explained by the Humean belief–desire model. Since anti-Humean motivations are not necessary to explain such action, Morgan’s Canon delivers yet another reason to reject anti-Humeanism. Now the argument will take a more pointed turn. It will be shown that anti-Humeanism should be abandoned outright, because it is incompatible with observations drawn from recent work in neuroscience. On the other hand, the mechanisms which actually operate to produce both moral and non-moral action *only* fit a Humean mould.

Studies in neurobiology and neuroeconomics have attempted to isolate the brain mechanisms responsible for moral action. For instance, a large body of evidence shows that, in animal brains, the ventral striatum and nucleus accumbens, along with the insula and the orbitofrontal cortex, respond to the satisfaction of basic biological needs such as food, shelter, and social contact. In humans, the same brain areas respond to these same goods, *and* to abstract rewards like receiving money. Furthermore one study by Moll and colleagues reports that increased activity in the ventral striatum is observed with fMRI scans when human subjects anonymously and voluntarily choose to donate money to a charity (Moll et al. 2006). In the experiment, subjects had the choice of donating up to US$128 to a real charitable organization, or not doing so. The ventral striatum was activated both by the receipt of pure monetary rewards and by decisions to donate. Indeed, the ventral striatum was *more* active when subjects made donations than when they received monetary rewards (Moll et al. 2006, p. 15624). Since the voluntary choice of donating to charity often issues from the moral judgment that donating is *the right thing to do*, such a choice can be regarded as a moral action. Moll et al.'s findings suggest that the same mechanisms which respond to the satisfaction of non-moral desires (e.g., for attaining food and money) also respond to the performance of moral action.

In another neuroimaging study, researchers compared human subjects' neural reactions to receiving money and being forced to give money to charity (Harbaugh et al. 2007). In a first treatment, the researchers forced subjects to donate money to a charity in a tax-like condition, and measured consequent increases in brain activity in the ventral striatum, the head of the caudate, and the nucleus accumbens. In a second treatment, the researchers measured increased activity in the same brain areas when subjects were *given* a sum of money. The authors discovered that the difference between these two measures reliably predicts people's willingness to donate to charity on a *voluntary* basis. People who exhibited greater neural responses (in the mentioned brain regions) to mandatory donations, relative to receiving money, were more willing to donate voluntarily (Mayr et al. 2009, p. 308).

Harbaugh et al. also suggest that activity in the ventral striatum and other brain regions is a neural indicator of the *utility* a subject receives both from getting money and from voluntarily giving money to charity (Harbaugh et al. 2007, pp. 1623–1624; Mayr et al. pp. 308–309). 'Utility' is a technical term behavioural scientists use to refer to the satisfaction of preference or desire. The authors support this conjecture by using measured increases in neural activity in the ventral striatum and insula as a basis for comparing the relative strength of an individual's preference for receiving money as compared to his or her preference for donating money. A selfish person who strongly prefers getting money to donating money would only be willing to give up very little of his money in order to donate an additional unit of it. But an altruistic person who has a stronger preference for donating money would be willing to give up considerably more of his money in order to donate. Harbaugh and colleagues found that by modeling their subjects as *utility-maximizers* in this way, they could accurately predict how much money each subject was willing to donate voluntarily to charity. It's important to appreciate that the theory of utility-maximization is itself a version of motivational Humeanism. Utility-maximizing agents are agents who act to maximize the satisfaction of their preferences or desires. Accordingly, Harbaugh's findings provide compelling empirical evidence in favour of Humeanism, because they show that a belief–desire model can successfully *predict* moral action.

At the same time, the neuroeconomics research makes trouble for anti-Humeanism. The most prominent arguments for anti-Humean motivations emphasize that these motivations are needed to explain *moral* thought and action specifically. Thus, for example, David McNaughton argues that besires must be postulated in order to explain how one can be motivated by a 'purely cognitive' awareness of a moral requirement:

> To be aware of a moral requirement is, according to the realist, to have a conception of a situation as demanding a response. Yet to conceive of a situation as demanding a response, as requiring one to do something, is to be in a state whose direction of fit is: the world must fit this state . . . .But the realist also wishes to insist that the agent's conception of the situation is purely cognitive. That is, the agent has a belief that he is morally required to act and so his state must have the direction of fit: this state must fit the world . . . . (McNaughton 1988, p. 109)

In McNaughton's view, besires are needed to account for how a moral judgment can motivate action, *given* the hypothesis that awareness of a moral requirement is a 'purely cognitive' state. Additionally, Little notes that virtue ethicists like McDowell (1998) invoke the notion of a desire-entailing belief for a very similar purpose:

> The virtue theorist's claim is that a kind of cognitive state – a kind of state that does satisfy a belief direction of fit – necessarily brings with it the motivation to act as it says we ought. There are certain *ways* of seeing or of conceiving the world, as many have put it, that one cannot have without reacting affectively in a certain way . . . . (Little 1997, p. 261)

Thus, from an anti-Humean perspective, when someone acts to meet the requirements of morality, i.e. whenever someone acts morally, he or she will be acting on a special type of cognitive state that both apprehends the morally relevant features of a situation and generates a motivation. Moreover, these anti-Humean motivations *only* motivate moral action. They are not necessary to explain non-moral actions like tying one's shoe, because these actions are not (necessarily) motivated by the recognition of any moral requirement.

Now if anti-Humean motivations are realized in brain processes, we should expect to see something *special* going on in people's brains when they undertake moral actions – something that would *not* be going on when they perform non-moral acts. If there were besires or desire-entailing beliefs, they would have special neural correlates, and those neural correlates would be at work when anti-Humeans say they are at work – viz. when people act in conformity with their moral judgments.

However, neural imaging of what happens in people's brains when they donate to charity – an act that can be construed as a moral act – have so far shown that *nothing* special is going on in the brain when people engage in moral action. Instead, the same neural mechanisms that are activated when people (*and* animals, for that matter) receive rewards, eat food, find shelter, or partake in social bonding are also at work when test subjects make morally significant charitable donations. By contrast, Humeanism is not in any way threatened by Harbaugh et al.'s conclusion that people give money to charitable organizations because 'these transfers are associated with neural activation similar to that which comes from receiving money for oneself' (Harbaugh et al. 2007, p. 1624). Humeans regard the motivations driving moral action to be of the same kind as non-moral (intentional) motivations: they are desire-like or affective states. So it is to be expected, from a Humean viewpoint, that the same brain processes which underlie motivations to acquire food and money would also underlie motivations to act morally. Furthermore, when activity in the striatum was modelled as an indicator of the utility one derives for charitable donations, donating behaviour was successfully predicted. The model assumes that people are motivated to voluntarily donate to charity because it is rewarding to them. The neural reward for voluntary giving is registered in the same way as the neural reward for getting money, since both events elicit activity in the ventral striatum. And yet, receiving money does not call for recognizing the moral requirements of a situation. So, the motivation to donate to charity stems from a neural reward that occurs independently of any state of mind which involves recognizing the moral requirements of a situation. The independence of the motivation-inducing reward for donating from any moral judgment about the rightness of donating is compatible with Humeanism, but quite contrary to anti-Humeanism (Zangwill 2008a, b).

8.6 Conclusion

Three considerations have been adduced as evidence in favour of the Humean theory of motivation, and against motivational anti-Humeanism. First, Humeanism is much more compatible with De Waal's theory of how motivations to act morally could have evolved from simpler precursors resembling the proto-moral motivations of nonhuman primates. Second, anti-Humean motivations are not uniquely necessary to explain any behaviour that could not be explained by the Humean belief–desire model. Indeed, anti-Humean motivations aren't even necessary to explain *moral* behaviour. And third, despite the anti-Humeans' insistence that only besires or desire-entailing beliefs can explain motivations to act in accordance with the recognition of moral requirements, neuroimaging studies of people engaging in moral action yield no indication of any special neural process which is not successfully explained by a Humean framework of utility maximization.

Humeanism carries significant implications for moral philosophy. It suggests that moral motivation cannot be a purely cognitive achievement. For even if there were mind-independent moral truths, knowledge of these truths wouldn't be sufficient to direct behaviour. People can't simply know how the world ought to be; they must also *want* to change the world accordingly. The study of how to increase people's desires to do justice is no less a worthy enterprise than the study of what justice is.

Acknowledgments Many thanks to David Wong, an anonymous referee, and the participants of the "What Makes Us Moral?" Conference, VU University Amsterdam, for their helpful feedback on this essay.

References

Altham, J.E.J. 1986. The legacy of emotivism. In *Fact, science, and morality: Essays on A.J. Ayer's language, truth and logic*, ed. G. Macdonald and C. Wright, 275–288. Oxford: Basil Blackwell.

Blackburn, S. 1984. *Spreading the word*. Oxford: Oxford University Press.

Brosnan, S.F. 2010. Fairness and other-regarding preferences in nonhuman primates. In *Moral markets: The critical role of values in the economy*, ed. P.J. Za and M.C. Jensen, 77–104. Princeton: Princeton University Press.

De Waal, F. 2006. Morally evolved: Primate social instincts, human morality, and the rise and fall of 'veneer theory'. In *Primates and philosophers: How morality evolved*, ed. S. Macedo and J. Ober, 1–58. Princeton: Princeton University Press.

De Waal, F. 2008. Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology* 59: 290–300.

Dennett, D.C. 1995. *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon & Schuster.

Gibbard, A. 1990. *Wise choices, apt feelings: A theory of normative judgment*. Cambridge, MA: Harvard University Press.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr. 2007. Explaining altruistic behavior in humans. In *Oxford handbook of evolutionary psychology*, ed. R.I.M. Dunbar and L. Barrett, 605–619. New York: Oxford University Press.

Haidt, J. 2003. The moral emotions. In *Handbook of affective sciences*, ed. R.J. Davidson, K.R. Sherer, and H.H. Goldsmith, 852–870. Oxford: Oxford University Press.

Harbaugh, W.T., U. Mayr, and D.R. Burghart. 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316: 1622–1625.

Joyce, R. 2006. *The evolution of morality*. Cambridge, MA: MIT Press.

Little, M.O. 1997/2007. Virtue as knowledge: Objections from the philosophy of mind. In *Foundations of ethics*, ed. R. Shafer-Landau and T. Cuneo, 252–264. Malden: Blackwell Publishing Ltd.

Mackie, J.L. 1977. *Ethics: Inventing right and wrong*. London: Penguin.

Mayr, U., W.T. Harbaugh, and D. Tankersley. 2009. Neuroeconomics of charitable giving and philanthropy. In *Neuroeconomics: Decision making and the brain*, ed. P.F. Glimcher, C.F. Camerer, E. Fehr, and R.A. Poldrack, 303–320. London: Elsevier.

McDowell, J. 1998. *Mind, value, and reality*. Cambridge, MA: Harvard University Press.

McNaughton, D. 1988. *Moral vision: An introduction to ethics*. Oxford: Basil Blackwell.

Moll, J., F. Krueger, R. Zahn, M. Pardini, R. Oliveira-Souza, and J. Grafman. 2006. Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences* 103(623): 15623–15628.

Moll, J., R. Oliveira-Souza, R. Zahn, and J. Grafman. 2008. The cognitive neuroscience of moral emotions. In *Moral psychology*, vol. 3, ed. W. Sinnott-Armstrong, 1–18. Cambridge, MA: MIT Press.

Pillutla, M.M., and J.K. Murnighan. 1996. Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Decision Processes* 68(3): 208–224.

Prinz, J.J. 2004. *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.

Rosati, C.S. 2006. Moral motivation. In *Stanford encyclopedia of philosophy*, ed. E.N. Zalta. http://plato.stanford.edu/entries/moral-motivation/. Accessed 29 June 2012.

Sanfey, A.G., J.K. Rilling, J.A. Aronson, L.E. Nystrom, and J.D. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300: 1755–1758.

Shafer-Landau, R. 2010. *The fundamentals of ethics*. New York: Oxford University Press.

Smith, M. 1987. The Humean theory of motivation. *Mind* 96(381): 36–61.

Sober, E. 1998. Morgan's canon. In *The evolution of mind*, ed. D. Dellarosa Cummins and C. Allen, 224–242. Oxford: Oxford University Press.

Thompson, P. (ed.). 1995. *Issues in evolutionary ethics*. Albany: State University of New York Press.

Zangwill, N. 2003/2007. Externalist moral motivation. In *Foundations of ethics*, ed. R. Shafer-Landau and T. Cuneo, 243–251. Malden: Blackwell.

Zangwill, N. 2008a. Besires and the motivation debate. *Theoria* 74: 50–59.

Zangwill, N. 2008b. The indifference argument. *Philosophical Studies* 138: 91–124.