The Moral Psychology Handbook

John M. Doris and the Moral Psychology Research Group

Fiery Cushman, John M. Doris, Joshua D. Greene, Gilbert Harman, Daniel Kelly, Joshua Knobe, Edouard Machery, Ron Mallon, Kelby Mason, Victoria McGeer, Maria W. Merritt, Shaun Nichols, Joseph M. Paxton, Alexandra Plakias, Jesse J. Prinz, Erica Roedder, Adina L. Roskies, Timothy Schroeder, Walter Sinnott-Armstrong, Chandra Sekhar Sripada, Stephen Stich, Valerie Tiberius, Liane Young

OXFORD
UNIVERSITY PRESS

Contents

1

Evolution of Morality[1]

EDOUARD MACHERY AND RON MALLON

> Biology provides a broad source of information about humans that has no substitute. It clarifies long-standing paradoxes. It shows that some things have indeed been missing from the debates about morality, and that they have been missing because the process of organic evolution that gave rise to all forms of life has been left out of the discussions.
>
> (Alexander, 1987: xvii)

Walking in Darwin's footsteps, numerous philosophers, psychologists, anthropologists, and biologists have turned toward evolutionary theory to provide a scientific understanding of morality.[2] In spite of their differences, these thinkers concur on the provocative claim that morality is an evolved part of human nature, much like a tendency to weave nets is an evolved part of spiders' nature.

This claim is supposed to have far-reaching implications in moral philosophy (e.g. Gibbard, 1990; D'Arms, ms). Proponents of evolutionary ethics have often attempted to justify specific moral norms by appealing to the evolution of morality (e.g. Spencer, 1892; Richards, 1986, 1989; Rottschaefer & Martinsen, 1990; Rottschaefer, 1991, 1998; Casebeer, 2003a).[3] The claim that morality evolved has also been used as a premise for various skeptical arguments about

[1] We would like to thank Steve Downes, Richard Joyce, Stefan Linquist, and particularly John Doris and Stephen Stich for their comments on previous versions of this chapter.

[2] Darwin (1871); Kropotkin (1902); Huxley (1894/1989); Waddington (1942); Trivers (1971); Singer (1981, 2000); Boehm (1982, 1999); Alexander (1987); Ruse & Wilson (1985); Frank (1988); Gibbard (1990); Irons (1991); Fiske (1991); Wright (1994); Cronk (1994); Dennett (1995); Kitcher (1998, 2006a, b); Wilson (2002); Levy (2004); Allen & Bekoff (2005); Krebs (2005); Joyce (2006); Hauser (2006). Lahti (2003) and Prinz (2008: ch. 7) are more skeptical.

[3] Evolutionary ethics is a specific philosophical tradition. In spite of their diversity, evolutionary ethicists concur that evolutionary theory leads to specific conclusions in normative ethics (i.e. that some particular moral norms are justified or unjustified). For a historical overview of this philosophical tradition, see Farber (1994).

morality (Ruse, 1986; Woolcock, 2000; Joyce, 2000, 2006; Street, 2006, 2008; for critical discussion, see, e.g., Sober, 1994; Copp, 2008).

While it matters philosophically whether or not morality is a product of evolution, we find ourselves agreeing with Darwall, Gibbard, and Railton's complaint that "more careful and empirically informed work on the nature or history or function of morality is needed (. . .) [p]erhaps unsurprisingly, very little such work has been done even by some of those who have recommended it most firmly" (1992: 34). Fifteen years after they expressed this complaint in their well-known article "Toward fin de siècle ethics: Some trends," it remains unclear whether, and in which sense, morality evolved. Our goal in this chapter is to answer these questions. Specifically, we propose to clarify the claim that morality evolved by distinguishing three possible versions of this claim and to review the evidence in support of each. We conclude that two versions of the claim that morality evolved are relatively well supported, but that they are unlikely to yield significant philosophical payoffs, while the stronger version, which is of real interest to philosophers, is in fact empirically unsupported.

Here is how we proceed. In Section 1, we examine a First interpretation of the claim that morality evolved—one on which *some components of moral psychology* have evolved. We argue that this claim is uncontroversial although it can be very difficult to show that some particular components of moral psychology really evolved. In Section 2, we turn to a second interpretation of the claim that morality evolved, the claim that *normative cognition*—that is, the capacity to grasp norms and to make normative judgments—is a product of evolution. We argue that normative cognition might well have evolved, and that it may even be an adaptation. Finally, we turn to the philosophically most interesting interpretation of the claim that morality evolved. In Section 3, we set out the view that *moral cognition*, understood as a special sort of normative cognition, is the product of evolution, and we argue that the evidence adduced in support of the view is unpersuasive.[4] We conclude by expressing our skepticism about the philosophical implications that can be drawn from the literature on the evolution of morality.

[4] We do not tackle here a whole range of issues that are often associated with the topic "evolution and morality." In particular, we do not discuss the philosophical tradition of evolutionary ethics, and we only indirectly examine the meta-ethical implications of the evolution of morality.

1. The Evolution of Components of Moral Psychology

1.1. *The Project*

As noted in the introduction, the claim that morality evolved can be interpreted in at least three different ways. The first interpretation asserts that specific components (e.g. emotions, dispositions, rule-based reasoning systems, or concepts) of moral psychology or specific behaviors typically associated with morality evolved. Some evolutionary theorists ask whether some of these components or behaviors evolved, whether they are adaptations, how they could have contributed to fitness, and whether they evolved exclusively in the hominid taxon or in other taxa.

Frans de Waal's work is a good illustration of this approach (e.g. de Waal, 1996; Preston & de Waal, 2002; see also Darwin, 1871; Bekoff, 2004). He is interested in whether some of the emotions, dispositions, and cognitive competences that underlie moral behaviors—e.g. empathy and the recognition of norms—are present in our closest extant relatives, the apes, as well as in more distant relatives, such as old-world and new-world monkeys. Thus, when he defines his project at the beginning of *Good Natured*, he asks, "Do animals show behavior that parallels the benevolence as well as the rules and regulations of human moral conduct? If so, what motivates them to act this way? And do they realize how their behaviors affect others?" (1996: 3).

This interpretation of the claim that morality evolved strikes us as not at all contentious although specific hypotheses about the evolution of particular components of moral psychology may be controversial. It is highly plausible that some moral emotions have an evolutionary history because many emotions have a long evolutionary history (e.g. Fessler & Haley, 2003). And the cognitive architecture of morality also relies on various components of social cognition, many of which also have a long evolutionary history (e.g. Fessler, 1999; Stone, 2006).

Although the idea is fairly uncontroversial, showing that a specific component of moral psychology evolved is difficult. In the remainder of this section, we focus on what is perhaps the main difficulty: before looking for

the homologues[5] of human moral traits[6] in other species, such as chimpanzees, researchers should establish that these traits are good candidates for being evolved traits.

1.2. *Fairness in Non-Human Primates?*

De Waal has long argued that many important components of moral psychology, such as the sense of fairness and numerous fairness-related emotions, e.g. gratitude (Brosnan & de Waal, 2002) and inequity aversion (Brosnan & de Waal, 2003; Brosnan, 2006), are homologous to psychological systems in other primates.[7] Here, we focus critically on de Waal's claim that there is evidence for a precursor of the human sense of fairness among female brown capuchins (Brosnan & de Waal, 2003; for related results with chimpanzees, see Brosnan, Schiff, & de Waal, 2005; and for dogs, see Rangea, Horna, Viranyi, & Hubera, 2009). Our goal is not to challenge the idea that many components of our moral psychology (psychological systems, emotions, etc.) evolved: as noted above, this claim strikes us as non-controversial. Rather, focusing on the example of the sense of fairness, our goal is to illustrate how difficult it is to show that some particular component evolved because some traits that might seem to be good candidates for being evolved traits might, on further examination, turn out to be poor ones.

Brosnan and de Waal's experimental design is clever. Capuchins, which have been trained to exchange coins for foods, are put in two adjacent cages. They are given a coin and have to give it back in order to receive a piece of food, which is visible in a transparent bowl in front of them. In one condition, the two capuchins are given a similar recompense, a piece of cucumber. In a second condition, one monkey receives a piece of cucumber, while the second monkey receives a piece of grape (a highly valued food). In a third condition, one monkey receives a piece of cucumber, while the second monkey is given a piece of grape without having to exchange it for a coin. Brosnan and de Waal measure the rate of rejection by monkeys, i.e. the number of cases where the monkeys do not exchange the coin or throw it. The results are surprising:

[5] As a first approximation, two traits are homologues if they are modifications of a single ancestor trait (see, e.g., the human eye and the chimpanzee eye) or if one is the modification of the other (see, e.g., the human eye and the eye of humans' and chimpanzees' last common ancestor) (for discussion, see Brigandt, 2003; Griffiths, 2006). Homologues are not necessarily very similar: for instance, mammals' arms and bats' wings are homologous, although they look quite different (at least superficially).

[6] "Trait" is a term of art in evolutionary biology. It refers to the physiological, behavioral, psychological, etc. properties of organisms (e.g. bipedality or a specific skull structure). This use is obviously different from the use of "trait" in the controversy about character in psychology and in ethics.

[7] De Waal (1996); Flack & de Waal (2000). See also Trivers (1971); Bekoff (2001, 2004).

female capuchins reject at a much higher rate the piece of cucumber when the other capuchin is given a grape for a coin and at an even higher rate when the other capuchin is given a grape for free.[8]

Brosnan and de Waal argue that this is tentative evidence for expectations about fair distributions of food, that is, for norms of fair distribution, as well as evidence for social emotions similar and homologous to human moral outrage. They write (2003: 299):

> People judge fairness based both on the distribution of gains and on the possible alternatives to a given outcome. Capuchin monkeys, too, seem to measure reward in relative terms, comparing their own rewards with those available, and their own efforts with those of others. They respond negatively to previously acceptable rewards if a partner gets a better deal. Although our data cannot elucidate the precise motivations underlying these responses, one possibility is that monkeys, similarly to humans, are guided by social emotions. These emotions, known as 'passions' by economists, guide human reactions to the efforts, gains, losses and attitudes of others.

And, in a related paper (2004: 140), they add:

> [C]apuchin monkeys react negatively when another individual gets a better reward for the same or less effort on a specific task. This finding suggests that precursors to inequity aversion are present in animals from which our lineage split millions of years ago.

We are skeptical, and we now argue that it is unlikely that capuchins obey a norm of fair distribution of windfall gains that is homologous with any human fairness norm. Let us emphasize that we are not denying that the sense of fairness—the tendency to find some actions fair and others unfair—has plausibly evolved. Our claim is more specific: we question whether Brosnan and de Waal's work provides evidence that a specific norm of fairness—the equal distribution of windfall profits—is a homologue present among capuchins and humans. We then use the example of Brosnan and de Waal's research to draw some cautionary conclusions about the search for homologues of the components of human morality.

First, Brosnan and de Waal (2003) found no effect for male capuchins (but see van Wolkenten et al., 2007). This is curious if Brosnan and de Waal have really identified a homologue of a human norm of fair distribution of windfall gains. Among humans, there is some variation in how males and females

[8] A control condition ensures that this result is not a mere effect of the presence of highly valued food. Note however that Brosnan, Freeman, & de Waal (2006) failed to replicate capuchin monkeys' aversion to inequity in a different experimental design, and that Bräuer, Call, & Tomasello (2006) failed to replicate chimpanzees' aversion to inequity. Brosnan and de Waal's design has also been severely criticized (Dubreuil, Gentile, & Visalberghi, 2006; but see van Wolkenten, Brosnan, & de Waal, 2007). We will bracket these issues here.

behave in similar situations (e.g. Andreoni & Vesterlund, 2001; Solnick, 2001). However, in an economic game called "the dictator game," both males and females are disposed to reject low offers (Solnick, 2001), suggesting that both get upset when windfall gains are shared unequally.[9]

In addition, Henrich (2004) has noted two problems with Brosnan and de Waal's proposal (but see Brosnan & de Waal's [2004] reply). First, in similar conditions, humans tend to react very differently from female capuchins. When they are offered a deal that they judge to be unfair, humans in many cultures reject this deal, when such a rejection hurts the person who offered the deal (Henrich et al., 2004). However, when rejecting the deal does not hurt the person who offered it, which is a situation analogous to the second and third conditions in Brosnan and de Waal's experiment, people tend to accept the deal, in sharp contrast with capuchins (Bolton & Zwick, 1995).

One could argue on behalf of Brosnan and de Waal that Henrich's first objection is unconvincing. Henrich is certainly correct that humans do not behave similarly to capuchin monkeys. However, it is plausible that in situations that are analogous to Brosnan and de Waal's experiments, humans in many cultures feel annoyed and angry. But because they are able to control their anger and to act in their best interest, humans accept the offer, when rejecting the offer would not hurt its author. By contrast, capuchins are not able to control their anger and are thus unable to act in their best interest. It would be easy to test the hypothesis that in situations that are similar to Brosnan and de Waal's conditions 2 and 3, humans and capuchins react similarly in that they both feel anger. In particular, focusing on humans, one could examine whether there is any facial micro-expression of anger (micro-expressions are facial expressions of emotions that last only a fraction of second because the agent tries to suppress or control her emotion). One could also examine whether the brain areas involved in negative emotions (particularly the insula) and in executive control (the dorsolateral prefrontal cortex and the anterior cingulate cortex) are activated in these situations.

More troubling is Henrich's second criticism. Henrich and colleagues have documented that there is much cross-cultural normative diversity in the norms bearing on the distribution of windfall gains (Henrich et al., 2004, 2005). For instance, Americans believe that a fair distribution of such gains consists in splitting them equally. By contrast, in a few small-scale societies, such

[9] A "dictator game" is one in which a windfall is divided by one person ("the dictator"), and the resulting distribution can be accepted or rejected by the other. If it is rejected, the two persons get nothing. Because even a small share of a windfall is better than nothing, economic rationality suggests that parties should accept even small shares rather than reject an unfair distribution.

as the Machiguengas of the Peruvian Amazon, people seem to expect the beneficiaries of windfall gains to keep the gain for themselves.

Of course, by itself, variation, including cultural variation, does not show that a trait (i.e., in the present case, the norm of splitting windfall gains equally) has not evolved. To begin with, different adaptations can be selected for in different human populations. Moreover, a trait can also be designed so as to take different forms in different environments, including different social environments (see, e.g., Draper & Belsky, 1990). Finally, a given adaptation often varies across environments because the environment in which organisms develop influences its development.

However, in this specific case, cultural variation suggests that the norm about the fair allocation of windfall gains was not selected for. If this norm is really present in capuchins (as de Waal and colleagues would have it), then it is very ancient: it had already evolved 30 million years ago, before the platyrrhines (the phylum to which capuchins belong) and the catarrhines (the phylum to which humans belong) split. If it is that ancient, then it should plausibly be species-typical, exactly like vision is. However, the cross-cultural research suggests that it varies across cultures, undermining the hypothesis that the norm of splitting windfall gains equally is an evolved trait that is homologous in capuchins and humans. Rather than being a trait homologous to old-world monkeys and humans, what is fair in the kind of situations considered by Brosnan and de Waal or by Henrich and colleagues is determined by the culture-specific norms governing economic interactions.

Henrich's second comment illustrates what is maybe the most important difficulty that accompanies attempts to discover homologues of human moral traits. Suppose that one is interested, as de Waal or Bekoff are, in finding homologues of some of the traits (emotions, norms, concepts, etc.) that constitute human moral cognition (and not in establishing that these traits are themselves evolved). Because two traits are homologues only if they evolved from a common ancestor trait, such a research project assumes that the relevant components of human moral cognition have evolved or, at least, that they are good candidates for being evolved traits. Thus, before looking for homologues of a given component of human moral cognition, it would seem important to ensure that there are no strong reasons to doubt that this component really evolved.[10] The existence of a trait in only a few cultures, its emergence in

[10] If one's interest does not lie in finding homologues of the components of human moral cognition, but in showing that these components themselves evolved, then it is appropriate to look for plausible homologues even if there are some reasons to doubt that the relevant components really evolved. For finding plausible homologues of some components of human moral cognition would provide very strong evidence that these components evolved.

some recent, historical times, or its acquisition by means of a domain-general learning mechanism are strong reasons to doubt that this trait evolved. Thus the cross-cultural variation of how windfall gains should be shared suggests that the norm of splitting windfall gains equally is unlikely to be an evolved trait. It is then pointless to look for homologues of this norm.

1.3. *Summary: The Evolution of Psychological Components of Moral Psychology*

Researchers often focus on some components of moral psychology, such as the norm of fairness. Then they attempt to determine the evolutionary history of these components, by studying whether other species, particularly other primates, also possess the relevant traits. We have argued that this first interpretation is uncontroversial: some, and perhaps many, components of moral psychology evolved. At the same time, hypotheses about the evolution of specific components are difficult to establish, in part because some traits that might seem to be good candidates for having evolved may, on further examination, turn out to be poor candidates. Looking for homologues of the components of moral psychology requires careful attention to a range of data from multiple fields of scientific inquiry to ensure that there are no strong reasons to doubt that these components evolved. Cultural psychology and anthropology are needed to establish that this trait is not culturally local while developmental psychology is needed to show that it is not acquired by means of a domain-general learning mechanism. In this section, we illustrated this difficulty by discussing Brosnan and de Waal's claim to have found a homologue of the human fairness norm about windfall gains.

As noted in the introduction, the claim that morality evolved is often supposed to have far-reaching implications in moral philosophy, but little attention has been dedicated to examining whether and in which sense morality evolved. Now, we have just argued that, under at least one interpretation, it is uncontroversial that it did: some components of moral cognition evolved. So, one might ask, what follows from the evolution of morality in this sense? As we shall now see, very little.

It is important to distinguish three strategies for answering this question. First, one might ask whether anything of interest in moral philosophy follows from the claim that some components of moral cognition evolved. We believe that the answer is probably negative since we do not see how the argument would go, and indeed we know of no philosopher who argued to significant philosophical conclusion from this premise. Second, one might attempt to derive moral conclusions from the evolution of *specific* components of moral cognition. For instance, D'Arms (ms) argues that research on the evolution

of "self-righteous anger"—the anger directed at people who are not angered by others' moral violations—has moral consequences. Arguments of this kind can take one of the following two forms. First, one could propose to derive moral norms about dispositions to act, character traits, etc. from facts about the functions of these dispositions and character traits.[11] For instance, Casebeer contends that "moral facts are reducible to functional facts" (2003b: 67). Although we do not have the space to discuss this first kind of argument at length, we are not sanguine about it. Although some normative propositions might be reducible to functional propositions—e.g. the claim that an organ is working as it should might be reducible to the claim that it fulfilled its function—we doubt that this can be done from moral propositions without falling prey to some version of the open question argument (see Joyce, 2006 for further criticism of Casebeer). The second type of argument is illustrated by D'Arms's discussion of the normative consequences of the evolution of self-righteous anger. D'Arms correctly notes that research on the evolution of a morally relevant trait can improve our knowledge about what this trait is or what it does. For instance, research on the evolution of self-righteous anger improves our understanding of the effect of self-righteous anger on the social stability of norms. And what a trait is or does is surely relevant to whether one should morally have this trait. Although this kind of argument is the most promising way of deriving moral consequences from some evolutionary findings about a specific component of morality, it is noteworthy that these consequences are not derived from the fact that this component evolved, but rather from what it is or what it does. So, just like the two strategies discussed above, this argumentative strategy does not establish that moral consequences follow from the evolution of the components of moral cognition.

2. The Evolution of Normative Cognition

We now turn to the second interpretation of the claim that morality evolved. Researchers interested in this second interpretation focus on normative cognition in general: they contend that normative cognition evolved (and often, that it is an adaptation). In this section, we explain this claim in more detail, and we argue that there is a small, but suggestive, body of evidence that normative cognition is an adaptation.

[11] In this context, functions are understood etiologically: roughly, *y* is the function of *x* if and only if the fact that *x* does *y* explains why *x* exists. For instance, the function of shame is to motivate people to apologize for having broken some norms if shame was selected for during evolution (and as a result, exists nowadays) because of this effect.

2.1. *Normative Cognition*

Although the nature of norms is a controversial topic in the social sciences (e.g. McAdams, 1997), we offer an informal account that should be acceptable to many social scientists. As we shall understand them, norms are attitudes toward types of actions, emotions, thoughts, or other traits. These norms are typically shared by many members of a given group and regulate people's behaviors, thoughts, emotions, characters, and so on. Their content essentially involves deontic concepts, such as SHOULD or OUGHT. Such norms can prescribe or forbid a thought, behavior, or any other characteristic, and may be associated with a disposition to punish those individuals who do not comply with the norms.

Normative cognition is underwritten by a complex cognitive architecture. People learn and assimilate, explicitly and implicitly, numerous norms; they are motivated to comply with them; and they typically expect others to comply with them. Emotions are also a key component of this cognitive architecture. Several negative emotions are triggered by norm violations (Haidt, 2003; Fessler & Haley, 2003). Norm violators are likely to feel shame or guilt (depending on which emotion is emphasized in their culture).[12] Victims of norm violations and third parties are likely to feel anger or disgust toward norm violators. These emotions motivate behavior: the anticipation of feeling ashamed and guilty motivates avoiding the violation of norms, shame and guilt motivate reparative behavior, and anger motivates punishment (e.g. Fehr & Gächter, 2002; Haidt & Sabini, 2000). Disgust causes third parties to distance themselves from norm violators, which results in the loss of cooperative opportunities for the norm violators. Anticipatory fear of shame or guilt often motivates norm compliance (Fessler, 2007).[13] In addition to these negative emotions, positive emotions are caused by norm compliance. People feel elevation when others endure some cost to comply with certain norms (Haidt, 2003).

It is remarkable, however, there has been little systematic work on human normative cognition. One exception is Chandra Sripada and Stephen Stich's (2006) article. Sripada and Stich argue for the existence of two cognitive systems subserving the psychology of norms: an acquisition mechanism and an implementation mechanism. The function of the acquisition mechanism is to learn the norms that are prevalent in one's culture, while the function

[12] Research shows that in some cultures (e.g. Indonesia), people are more prone to feel shame than guilt when they violate a norm, while in other cultures (e.g. the USA), they are more prone to feel guilt than shame (Benedict, 1946; Fessler, 2004).

[13] For instance, according to J. Heinrich (personal communication, 10/21/2007), Fijians are constantly weighing the prospects of feeling shame when they make decisions.

of the implementation mechanism is to store representations of these norms, to produce some intrinsic desires to comply with them, and to motivate people to punish norm violators. While their hypothesis is consistent with the existence of innate representations of norms, Sripada and Stich speculate that the implementation mechanism does not store any innate representation of norms. Rather, children, and sometimes adults, need to learn the prevalent norms of their social community.

2.2. *How to Study the Evolution of Normative Cognition?*

Many researchers' work on the evolution of morality is best understood as being about the evolution of normative cognition in general, since they do not single out a specific kind of norms (i.e. moral norms).[14]

Before going any further, it is worth noting that there are many ways to investigate the evolution of a trait. It is particularly useful to distinguish two related claims. To claim that a trait *evolved* is simply to claim that the trait has a phylogenetic history, and one project would be to inquire into this history.[15] That is, one can study what changes took place in the psychology of our primate ancestors during the evolution of normative cognition (just as one can study the evolution of the human eye by identifying the changes that took place during the evolution of the mammalian eye). A stronger claim is that normative cognition constitutes an *adaptation*. An adaptation is a specific sort of evolved trait—i.e. a trait whose evolution is the result of natural selection. Since not all products of evolution are adaptations, someone who conjectures that normative cognition is an evolved trait can also examine whether it is an adaptation, the by-product of another adaptation, or an evolutionary accident. In addition, if one proposes that normative cognition is an adaptation, one should consider what its evolutionary function might be—that is, what selective forces might have driven its evolution.

2.3. *Evidence that Normative Cognition is an Adaptation*

Sociological and psychological evidence suggests that normative cognition is an adaptation. We consider these two types of evidence in turn (for further evidence, see Cummins, 1996b).

Norms, either informal or formal, are ancient: the historical record has no trace of a society without norms. Furthermore, norms are universal (although

[14] See, particularly, Fiske (1991); Bowles & Gintis (1998); Richerson, Boyd, & Henrich (2003); Gintis, Bowles, Boyd, & Fehr (2003); Nowak & Sigmund (2005).

[15] Phylogeny is the change of lineages through time. One looks at an evolved trait in a given species from a phylogenetic perspective when one considers how this trait results from changes to the traits possessed by the ancestor species of the species under consideration.

the content of norms varies tremendously across cultures). Small-scale societies are typically regulated by informal norms, while large-scale societies are typically regulated by informal and formal norms. All known societies also have policing mechanisms that ensure people's compliance with the prevalent norms (Brown, 1991). These policing mechanisms naturally vary across cultures. In some societies, but not in all, policing is the socially sanctioned role of a dedicated group of individuals (e.g. policemen, Iran's "moral" police [a branch of the Islamic Revolutionary Guard], etc.). In addition, in all societies, informal social practices contribute to ensure people's compliance with the prevalent norms (Boehm, 1999). These include gossip (Dunbar, 1996) and various forms of ostracism (Brown, 1991). Finally, as noted by Sripada and Stich (2006), norms permeate people's life: few behaviors and decisions are immune to the influence of some norm or other.

The antiquity and universality of norms is evidence that normative cognition evolved. When a trait is ancient and universal, it is either because it can be easily acquired by individual learning or by social learning, or because a developmental system is designed to ensure its regular development. In the latter case, but not in the former case, the universality and antiquity of a trait is evidence that it evolved. Ancient and universal traits that are *not* evolved, such as the belief that the sun rises every morning, are easy to acquire from one's physical and social environment (Dennett, 1995). Since it is difficult to see how one could acquire the capacity for normative attitudes toward thoughts, behaviors, and other traits—i.e. a capacity for norms—from one's environment (in contrast to acquiring specific norms, which can obviously be learned), it is plausible that normative cognition evolved.

Turning from sociological to psychological considerations, evidence suggests that people are endowed with a reasoning capacity that is specific to the domain of norms. While people reason poorly about non normative matters, they are adept at reasoning about normative matters (for review, see Cosmides & Tooby, 2005). Both Western and non-literate Shuar Amazonian subjects easily determine in which situations deontic conditionals, such as "If you eat mongongo nut (described as an aphrodisiac in the cover story), then you must have a tattoo on your chest" (described as a mark denoting married status), are violated, while they are surprisingly poor at determining in which situations indicative conditionals, such as ("If there is a red bird in the drawing on top, then there is an orange on the drawing below"), are false (Cosmides, 1989; Sugiyama, Tooby, & Cosmides, 2002). Although the interpretation of these findings remains somewhat controversial (e.g. Sperber, Cara, & Girotto, 1995), they suggest to us that people are distinctively adept at detecting norm violation.

Furthermore, just like adults, young children are much better at reasoning about the violations of deontic conditionals than about the falsity of indicative conditionals (Cummins, 1996a; Harris & Núñez, 1996). For instance, Cummins (1996a) showed 3-year-old children some toy mice and told them that some, but not all, could squeak. She also told them that some squeaky mice were inside the house, while others were outside. Finally, she told children that a cat was hunting mice outside the house, but only when they squeaked. Half of the children were told that Queen Minnie Mouse had told the mice, ''It's not safe outside for the squeaky mouse, so all squeaky mice *are* in the house.'' Those children were then asked to say which mice must be examined to see whether Minnie Mouse was right. The other half was told that Queen Minnie Mouse had told the mice, ''It's not safe outside for the squeaky mouse, so all squeaky mice *must stay* in the house.'' Those children were then asked to say which mice must be examined to see whether Minnie Mouse's rule has been broken. While almost 65% of 3-year-olds answered correctly the second question, only 30% of them answered correctly the first question. These findings suggest that the capacity to reason about norms develops early (as early as children's fourth year) and in a distinctive manner (since it seems independent from children's capacity to reason about conditionals in general).

The existence of a cognitive system that seems dedicated specifically to produce good reasoning about norms from an early age on provides some suggestive evidence that normative cognition is an adaptation.[16] Generally, the functional specificity of a trait is (defeasible) evidence that it is an adaptation. Furthermore, the fact that a trait develops early and that its development is distinctive—it is independent from the development of other traits—suggests that natural selection acted on its developmental pathway. The early development of a psychological trait suggests that it is not acquired as a result of our domain-general learning capacity; the distinctive development of a psychological trait suggests that it is not acquired as a by-product of the acquisition of another psychological capacity (for further discussion, see Machery, forthcoming). Thus, evidence tentatively suggests not only that normative cognition is an evolved trait, but also that it is an adaptation.

The findings just considered provide suggestive (though inconclusive) evidence that normative cognition is an adaptation. It is instructive to anticipate

[16] Note that this claim is independent of Cosmides and Tooby's more specific claims about the form the adaptation takes (e.g. Cosmides & Tooby, 2005). It might even be compatible with critiques of Cosmides and Tooby (Fodor, 2000) according to which differential reasoning about norm violations is due to the use of deontic concepts in the norms themselves (see Cosmides & Tooby, 2008a, 2008b, 2008c; Fodor, 2008; Mallon, 2008).

Section 3 and to compare these findings with the body of evidence typically adduced to support the claim that moral cognition, conceived as a specific kind of normative cognition, evolved. While researchers often claim that moral cognition, conceived as a specific kind of normative cognition, is universal, we shall argue in the next section that the evidence for this claim is lacking. By contrast, the evidence for the antiquity and universality of norms is extremely solid. We shall also challenge the claim that a key component of moral cognition—i.e. grasping the distinction between properly moral norms and conventional norms—develops early and reliably. By contrast, although the research on adults' and children's capacity to reason with deontic conditionals is not entirely uncontroversial, it is on safer ground. In this case, psychologists have indeed typically not challenged the claim that people reason better with deontic conditionals than with indicative conditions; rather, they have focused on how this difference is to be explained.

2.4. *"How-Possible" Models of the Selection of Normative Cognition*

In addition to this small, but suggestive, body of evidence that normative cognition in general is an adaptation, several models show how normative cognition *could* have been selected for during the evolution of hominids.[17] These "how-possible" models (Brandon, 1990) do not establish how normative cognition actually evolved: evidence is lacking to answer this question. But these models show, first, that the hypothesis that normative cognition was selected for is consistent with our knowledge of evolution; second, the selection of normative cognition in evolutionary models is robust: in several possible evolutionary situations—those represented by the how-possible models—normative cognition would have been selected for.[18]

Since the 1980s, Robert Boyd, Peter Richerson, and their colleagues have developed a series of models explaining how norms can be stable in a community. Here, we present two of their models informally. In a well-known model, Boyd and Richerson (1992) have shown that punishment (actions inflicting a cost on norm violators) can stabilize any norm, including norms that prescribe costly behaviors such as cooperation. Suppose for an instant that punishment is cost-free—the punisher does not endure any cost when she punishes. By violating a norm, norm violators might get some

[17] By contrast, the models that account for the evolution of morality, understood as a specific form of normative cognition, are not particularly plausible (see Section 3).

[18] Boyd & Richerson (1992); Henrich & Boyd (2001); Boyd et al. (2003); Gintis et al. (2003); Richerson et al. (2003); Richerson & Boyd (2005); Boyd & Mathew (2007); Hauert et al. (2007).

benefit or might avoid some cost, when norm compliance is costly. However, because they are punished, violators suffer a cost and do less well than those who comply with norms, but avoid punishing (''lazy norm compliers'') and those who comply with norms and enforce them (''punishers''). If successful behaviors tend to become common in a population (maybe because they are imitated by others), then compliance with norms will prevail. Thus punishment can stabilize norms. Importantly, in this model, norm compliance does not depend on the content of the norms, only on the punishment of norm violators. Thus different norms might be stabilized in different societies, consistent with the diversity of norms across cultures.

But, of course, punishment is not cost-free, although, in humans, it might be low-cost because of the development of weapons and of language (which allows people to hurt others by gossiping negatively about them). Because punishment is not cost-free, lazy norm compliers do better than punishers. Thus compliance without punishment might become more common in a population at the expense of compliance with punishment (Boyd & Richerson, 1992). However, if lazy norm compliers become more common, norm violators will in turn increase in frequency, because they will be less often punished. This will prevent the stabilization of norms. How, then, is norm compliance obtained?

This problem has been addressed in various ways. One could first suggest that punishment itself is a norm, and that lazy norm compliers get punished when they fail to punish norm violators, a type of punishment called ''metapunishment'' or ''second-order punishment'' (Boyd & Richerson, 1992). However, this suggestion only pushes the problem one step further, because metapunishment is itself costly.

Henrich and Boyd (2001) have proposed an alternative solution (for a different model, see Boyd et al., 2003). In their model, behaviors are transmitted culturally, but biased by *conformism* and *prestige*. Conformist bias means that common behaviors are more likely to be transmitted than rare behaviors, while prestige bias means that high-payoff behaviors are more likely to be transmitted than low-payoff behaviors. Conformism favors the cultural transmission of the prevalent behaviors, whatever these are, while prestige-biased transmission can undermine norm compliance, punishment, and metapunishment, because these can be costly. Suppose now that, in a population, everybody complies with the norms, but fails to punish norm violators (everybody is *a lazy norm complier*). An intruder who fails to comply with the norms (*a norm violator*) would be better off than these lazy norm compliers, and prestige bias would tend to lead others to become norm violators, providing that this bias was stronger than the countervailing bias to conform with the more common

compliant behavior. So, where conformism is weak, norm violation will become common.

Consider now a second case. Everybody complies with the norms and punishes violators (everybody is *a punisher*). An intruder norm violator would not be better off than the common punishers, because she would be punished. But an intruder *lazy norm complier* would be better off than the common punishers, since she would not get punished (since she complies with the norms) and would avoid the cost of punishing others. By contrast, punishers would pay the cost of punishing the other punishers who would fail by mistake to comply with the prevalent norms. The extent of a lazy norm complier's advantage over the punishers depends on how costly it is to punish and on how often punishers fail to comply by accident with the prevalent norms.[19] If this advantage is large enough to offset the advantage conformism gives to punishers' common behavior (that is, to offset the fact that due to people's conformism, common behaviors are more likely to be imitated by others than rare behavior), compliance with the norms without punishing would become common. If lazy norm compliers replace the punishers in a population, the norm violators will ultimately invade this population.

Now, consider a third case. Everybody complies with norms, punishes violators, and punishes non-punishers (everybody is *a metapunisher*). An intruder lazy norm complier would not be better off than the common metapunishers, because she would be punished for non-punishing the metapunishers' accidental norm violations. But an intruder who would comply with the prevalent norms, punish violators, but fail to punish those who fail to punish (*a lazy punisher*) would be better off than the common metapunishers because she would not be punished (she complies with the norms) and because she would avoid the cost of punishing the failure to punish. By contrast, metapunishers would pay the cost of punishing those metapunishers who would fail by mistake to punish the metapunishers who by mistake violate a norm. However, the advantage of a lazy punisher over the metapunishers is smaller than the advantage of a lazy norm complier over the punishers, for the former advantage depends on *two* mistakes, i.e. a metapunisher failing by accident to comply with a prevalent norm and another metapunisher failing by accident to punish the accidental non-compliance. The lazy punisher's advantage is thus less likely to offset the advantage conformism gives to the common metapunishing behavior. Of course, the same argument applies at further orders of punishment. Thus, even if conformism is weak, it can stabilize punishment at some order

[19] This punisher is a cooperator who fails to cooperate by accident. Think for instance of someone who was unable to fulfill her promise to pick up a friend at the airport because her car failed to start.

of punishment. If punishment is stable, then norm compliance is itself stable. Thus Henrich and Boyd show that with a small amount of conformism that stabilizes metapunishment (or some higher-order punishment), costly norm compliance is stable: compliance with the prevalent norms, even at one's own cost, is more likely to be culturally transmitted than non-compliance with these norms.

Now, suppose that, as this and other models suggest is possible, cultural transmission stabilized norms during human evolution. Because norm violators were punished for violating the prevalent norms and because lazy norm compliers were punished for not punishing norm violators, both norm violators and lazy norm compliers incurred costs that punishers (who comply with the norms and punish) did not incur. Our ancestors who were good at learning the prevalent norms and who were motivated to comply with them and to punish norm violators might thus have had a fitness advantage over people who learned badly the prevalent norms or had a weak (if any) motivation to comply with them. Thus natural selection might have favored some important elements of the architecture of normative cognition—a disposition to learn prevalent norms, a disposition to comply with norms, and a disposition to punish norm violators.

2.5. *Summary: The Evolution of Normativity*

In this section, we have focused on a second interpretation of the claim that morality evolved: normative cognition—the capacity to grasp and apply norms—evolved. A small body of evidence suggests that normative cognition evolved by natural selection. Furthermore, several how-possible models show how normative cognition could have been selected for during the evolution of the human species.

Importantly, this conclusion is cold comfort to those philosophers who want to get some philosophical mileage out of evolutionary findings. This is particularly clear when one focuses on the argument that the evolution of morality would undermine the authority of moral norms (e.g. Ruse, 1986; Joyce, 2006). Suppose that this argument from the evolution of morality is meant to hang on the reading of the claim that morality evolved considered in this section: normative cognition in general evolved. While this argument would then rest on a premise that is supported by a small, but convincing body of evidence, it would have very troubling consequences. If the evolution of normative cognition really undermines the authority of moral norms, then it should also undermine the authority of *any* kind of norms (including epistemic norms), for there is no reason why *only* the authority of moral norms would be undermined by the evolution of the capacity to grasp norms *tout court*.

The unpalatable nature of this conclusion would plausibly give grounds for concluding that the argument from the evolution of morality is flawed. The upshot should be clear: if the claim that the evolution of morality undermines the authority of morality is to be plausible at all, it has to be based on an interpretation of the claim that morality evolved different from the one considered in this section. In our view, it is no accident that when philosophers have attempted to derive philosophical implications from the hypothesis that morality evolved, they have typically focused on a third reading of this hypothesis. We now turn to this reading.

3. The Evolution of Moral Normativity

3.1. *The Project*

Researchers who endorse the third version of the claim that morality evolved start by drawing a distinction among different types of normative cognition and by singling out one specific type of normative cognition, which they call ''morality.'' They then proceed to argue for the evolution of this type of normative cognition. Consider each step of this project.

3.1.1. *Morality as a Type of Normativity* As we saw in Section 2, normative cognition includes the capacity to grasp norms, to make normative judgments, and to be motivated to act according to the norms that one endorses.[20] Norms have to do with the regulation of people's actions, emotions, thoughts, or other traits. They specify what kinds of behaviors, emotions, thoughts, or other characteristics are mandatory, permissible, or recommended.[21] In turn, normative judgments consist in judging that behaviors, emotions, and thoughts (one's own or others') are unconditionally mandatory, permissible, or recommended.[22]

Turn now to the claim that moral cognition is a distinctive type of normative cognition. The basic idea is that moral norms are a distinct type of

[20] We distinguish grasping a norm from making a normative judgment because research on psychopathy suggests that it is possible to grasp a norm without endorsing it (Roskies, 2003; but see Levy, 2007; Prinz, 2008, ch. 1).

[21] For a discussion of normativity, see, e.g., Gibbard (1990: 61–80) and Railton (1999).

[22] Note that saying that a behavior (emotion, etc.) is unconditionally mandatory (permissible, etc.) is not the same as saying that it is universally mandatory (permissible, etc.)—that is, that it is mandatory for everybody. One can make unconditional normative judgments that apply only to some groups of people. For instance, one could judge that some actions are permissible for adults, but not for children, that some actions are forbidden for some particular social groups, such as a caste, etc.

norm and that related entities like moral judgments, moral motivations, and moral behaviors and thoughts are similarly distinct. For the sake of simplicity, we focus our discussion here especially on norms and normative judgments. There are many kinds of normative judgments, and moral judgments are only one of them. Other kinds of normative judgments might include judgments about what is rational (e.g. "you shouldn't believe that, given what else you believe"), aesthetically appropriate ("one should never wear green pants with a yellow shirt"), prudent ("if you want to live a long life, you should wear your seatbelt"), and conventionally expected ("if you are satisfied with the service, the tip should be at least 20%"). The first interpretation of the claim that morality is the product of evolution rests on the idea that moral judgments provide a *distinctive* means of regulating or evaluating actions, emotions, intentions, or character.

Richard Joyce's (2006) book *The Evolution of Morality* provides a particularly clear illustration of the kind of research considered here (see also Ruse, 1986; D'Arms, 2000; Joyce, 2008a, b). Focusing on moral judgments, he proposes that seven properties distinguish moral judgments and moral behaviors from other kinds of normative judgments:

- Moral judgments (as public utterances) are often ways of expressing conative attitudes, such as approval, contempt, or, more generally, subscription to standards; moral judgments nevertheless also express beliefs; i.e., they are assertions.
- Moral judgments pertaining to action purport to be deliberative considerations irrespective of the interests/ends of those to whom they are directed; thus they are not pieces of prudential advice.
- Moral judgments purport to be inescapable; there is no "opting out."
- Moral judgments purport to transcend human conventions.
- Moral judgments centrally govern interpersonal relations; they seem designed to combat rampant individualism in particular.
- Moral judgments imply notions of desert and justice (a system of "punishments and rewards").
- For creatures like us, the emotion of guilt (or "a moral conscience") is an important mechanism for regulating one's moral conduct. (2006: 70–71)

He adds that "so long as a kind of value system satisfies *enough* of the above, then it counts as a moral system" (71).

While Joyce clearly intends his list to function not as a checklist of necessary and sufficient conditions, but rather as something like a cluster concept, it is worth emphasizing that his claim is substantive and provocative precisely because of the rich characterization of moral judgments that he offers. That is,

according to his account, moral judgments have many distinctive properties that differentiate them from other sorts of normative judgments. In expressing skepticism about whether the capacity to make moral judgments is a product of evolution, we mean specifically to doubt Joyce's view and others like it. We do not doubt that there exists some thin description of the class of moral judgments that could be offered such that, under this description, the capacity to make moral judgments would be the product of evolution.[23] We deny the claim that when moral judgments are richly described, the capacity to make them is a product of evolution.

3.1.2. *Morality as an Evolved Trait* After characterizing distinctively moral cognition—for example, after characterizing moral norms as a distinct kind of norm or moral judgments as a distinct kind of normative judgment—researchers conjecture that such distinctively moral cognition is an evolved trait. Remember that in Section 2 we distinguished two related ways of studying the evolution of a trait. One could simply claim that morality, as a specific kind of normative cognition, *evolved*. One would then study what changes took place in the psychology of our primate ancestors during the evolution of the grasp of distinctively moral norms and of the capacity to make moral judgments. Since, as noted in Section 2.2, not all products of evolution are adaptations, someone who conjectures that the capacity to grasp moral norms and the capacity to make moral judgments are evolved traits can also examine whether they constitute an adaptation (as Dennett, 1995, Kitcher, 1998, and Joyce, 2006 have claimed), whether it is a by-product of another adaptation, or whether it is an evolutionary accident (Williams, 1988). In addition, if one proposes that the grasp of moral norms and the capacity to make moral judgments constitute an adaptation, one should consider what its evolutionary function might be—that is, what selective forces might have driven its evolution.

Joyce, for example, suggests that the capacity to make moral judgments is a specifically human adaptation for motivating us to act in a prosocial way. In essence, moral judgments provided our ancestors with compelling reasons to act in ways that typically favor others and that can be detrimental to themselves (see also Dennett, 1995). As a result, moral judgments reliably caused prosocial behavior. Moreover, loosely following Robert Frank (1988), Joyce contends that because moral judgments can be linguistically expressed, they signal to

[23] Indeed, as we have seen in Section 2, we allow that normative cognition *tout court*—as opposed to distinctively moral normative cognition—may well be a product of evolution.

others that we are committed to act in a prosocial way. The capacity to make moral judgments was favored by natural selection because reliable prosocial behavior and the signaling of one's dispositions to act in a prosocial way were favored during the evolution of the human species, possibly because prosocial behaviors were reciprocated.

Joyce is not the only researcher to claim that the capacity to make moral judgments, understood as a distinct type of normative judgment, is an adaptation. In *Moral Minds*, Marc Hauser (2006) contends that, like the language faculty, the moral faculty is a distinct psychological adaptation, although he has little to say about its evolutionary function (2006: xvii):

> The central idea of this book is simple: we evolved a moral instinct, a capacity that naturally grows within each child, designed to generate rapid judgments about what is morally right or wrong based on an unconscious grammar of action.[24]

In contrast to these projects, we see little reason to believe that the grasp of distinctively moral norms and the capacity to make moral judgments, understood as a specific kind of normative judgments, evolved at all, and so we doubt that they constitute an adaptation, a by-product, or an evolutionary accident. We conjecture that in this respect, the capacity to grasp moral norms and the capacity to make moral judgments might be similar to chess or handwriting. The capacities to play chess and to write involve various evolved cognitive traits (e.g. visual recognition and memorization of rules for the former), but they did not evolve. Similarly, we conjecture that the capacity to grasp moral norms and the capacity to make moral judgments involve various evolved cognitive traits (including, as we proposed in Section 2, a disposition to grasp norms in general), but they themselves did not evolve. In any case, as we shall argue now, none of the available evidence suggests that they did.

In the remainder of Section 3, we look critically at two different forms of argument for the claim that moral cognition evolved:

(1) There are plausible adaptationist models that predict its selection. Thus the grasp of moral norms and the capacity to make moral judgments are likely to be an adaptation and, *a fortiori*, to have evolved.
(2) The universality and innateness of the capacity to grasp moral norms and to make moral judgments is evidence that it is an evolved trait.

[24] For other approaches to the evolution of moral cognition, understood as a distinct type of normative cognition, see Darwin (1871), Ruse & Wilson (1985), Ruse (1986), Dennett (1995: chs. 16–17), Kitcher (1998), Singer (2000), and Levy (2004).

3.2. *Adaptationist Models of the Evolution of Morality*

It is common to argue that a trait is an adaptation by showing that there are plausible adaptationist models that predict the selection of this trait. For instance, some sociobiologists have provided this kind of argument in support of the hypothesis that female orgasm is an adaptation: they have argued this because it is plausible that, among our ancestors, those females who were able to have orgasms were more motivated to have sex and, as a result, were more likely to have descendants than those females who had no orgasm (for review and criticism, see Lloyd, 2005). This kind of argument has been severely criticized in the philosophy of biology because it involves telling just-so stories that cannot be supported by any available evidence (Gould & Lewontin, 1979; Kitcher, 1985).[25] Here, we do not discuss the value of this kind of argument in general. Rather, we criticize specific adaptationist models of the evolution of morality. We argue that these models do not support the claim that moral cognition, conceived as a distinct kind of normative cognition, was selected for. We first consider models that appeal to *reciprocal altruism*, then we consider models that are based on *indirect reciprocity*, before finally raising a general problem for all current adaptationist models of the evolution of morality.

Many adaptationist models of the evolution of morality appeal to a specific evolutionary mechanism, reciprocal altruism (Figure 1.1). The notion of reciprocal altruism was developed by evolutionary biologist Robert Trivers as a possible explanation of altruistic behavior among non-related organisms (Trivers, 1971).[26] The idea goes roughly as follows: a gene G for an altruistic trait T is favored by natural selection if T benefits discriminatively recipients who are likely to reciprocate in the future. Reciprocation is delayed and may be of a different kind (as happens, e.g., when chimps exchange grooming for political support, a phenomenon reported in Foster et al., 2009). To use a toy example, a gene G for sharing food is favored by natural selection if the bearer of G shares food with individuals that are likely, at some point in the future, to reciprocate by acting in a way that increases the fitness of the bearer

[25] Gould and Lewontin illustrated just-so stories with sociobiologist David Barash's work on bluebirds (but see Alcock, 2001: 65–68). Having observed that male bluebirds attack significantly more stuffed males near their nets before than after the eggs were laid, Barash speculated that males' aggressiveness toward other males was an evolved disposition for avoiding cuckoldry. Although this hypothesis seems to make sense of the trait under consideration, Gould and Lewontin argued that it was simply an untestable speculation. Research on the evolution of human brain size or human language offers numerous other examples.

[26] In evolutionary biology, a trait is said to be altruistic if it reduces the individual relative fitness of the bearer of the trait (the cost of the trait for the bearer) while increasing the individual relative fitness of another individual (see Chapter 5 of this volume on altruism).

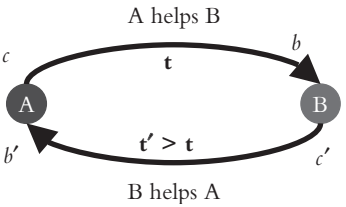

Figure 1.1. Reciprocal altruism
Note: c and c' stand for the costs of the altruistic actions for the agents A and B, b and b' for the benefits bestowed upon the beneficiaries of the actions, A and B ($c < b'$, $b > c'$), and t and t' for the times of the actions.

of G. According to Triver's model, three conditions have to be met for reciprocal altruism to explain the evolution of altruism. First, for the two individuals involved in a reciprocal interaction, the cost of being altruistic (in units of fitness) must be lower than the benefit received from the reciprocator's altruism. The cost of sharing food for the bearer of G must be lower than the benefit taken from the reciprocation at a later time. Second, the benefit of altruism must be withheld from those individuals who did not reciprocate in the past—usually called "cheaters." The bearer of G should refrain from sharing food with those individuals who did not reciprocate in the past. Third, individuals must interact repeatedly, so that the cost to cheaters of foregone protracted reciprocal interactions (in units of fitness) is greater than the benefit cheaters take from non-reciprocating. The bearer of G must share food with individuals with whom she is likely to interact often, so that not benefiting from food sharing over a long period of time is more costly than avoiding the cost of reciprocating a past benefit.

It is likely that reciprocal altruism fails to explain a range of human altruistic behaviors. As we saw, Trivers's idea is that if interactions between two individuals last for long enough, traits that benefit discriminatively those individuals that are likely to reciprocate will be favored by natural selection. This hypothesis fails to explain why people are disposed to benefit individuals with whom they will probably have no further interaction. For instance, people regularly tip waiters while on vacations, even though they will have no further interactions with them.

To explain away this difficulty, one might suggest that our ancestors lived in close-knit societies, where interactions were typically repeated. As a result, we evolved cooperative dispositions that do not distinguish between interactions that are likely and interactions that are unlikely to be repeated. In substance, we evolved to treat every interaction as if it was likely to be repeated (e.g. Johnson, Stopka, & Knights, 2003). This reply won't do, however (Fehr &

Henrich, 2003). Anecdotal reports and experimental evidence in behavioral economics show that people effortlessly distinguish between interactions that are unlikely to be repeated and long-term cooperative situations, and that they behave differently in these two types of situations. For instance, in experimental contexts, people tend to be less generous and helpful in the first type of situations than in the later type of situations (e.g. Gächter & Falk, 2002). Thus it does not seem to be the case that we behave cooperatively in some non-repeated interactions (e.g. when we tip strangers) because we evolved to treat every interaction as if it is likely to be repeated.[27]

Moreover, it is unclear whether our ancestors really lived in small and close-knit communities. Reciprocal altruism can explain the selection of altruistic behaviors only if our ancestors were able to discriminate between those individuals who failed to reciprocate altruistic acts ("cheaters") and those who did reciprocate, that is, only if they were able to remember who did what. The larger the group in which our ancestors belonged and the more fluid membership in residential units was,[28] the less likely it is that this condition was met. Paleoanthropological evidence suggests that at least for the last 50,000 years, our ancestors have lived in large groups of several thousands of members—too large for them to have been able to remember who did what (see Richerson & Boyd, 1998, 1999 for a detailed review of the evidence). Furthermore, as Richerson and Boyd write (1999: 254), "Foraging societies are simple by comparison with modern societies, but even the simplest contemporary hunting and gathering peoples, like !Kung San and the peoples of Central Australia, link residential units of a few tens of people to create societies of a few hundred to a few thousand people." Migrations and long-distance economic exchanges have also characterized the life of our ancestors for maybe several hundreds of thousands of years (McBrearty & Brooks, 2000). Finally, in many modern hunter-gatherer societies, membership in residential units is fluid (see, e.g., Hill, 2003 on the Ache; Smith, 2004 on the Hadza).

Clearly, these findings do not establish beyond doubt that reciprocal altruism could not explain the evolution of altruism. Morality could have evolved before our ancestors lived in large and fluid groups. Furthermore, even if our ancestors lived in large and fluid groups, they might have interacted altruistically only with a small number of group members. Nonetheless, the body of evidence about the size and fluidity of the social groups that have been common during

[27] One could object that people might have learned to override their tendency to behave altruistically, a tendency that could have been selected for by reciprocal altruism. We concede that this is a possibility, but we believe that some evidence would be required to substantiate this hypothesis.

[28] Membership is fluid when people can easily join and leave residential groups. When membership is fluid, people are more likely to interact with strangers.

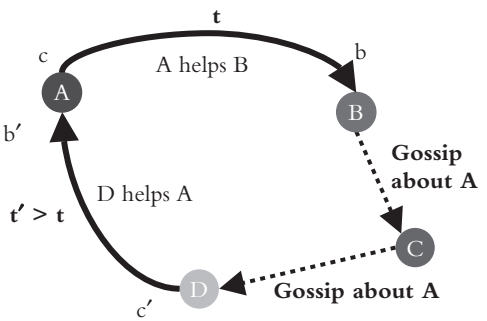

Figure 1.2. Indirect reciprocity
Note: c and c' stand for the costs of the altruistic actions for the agents A and D, b and b' for the benefits bestowed upon the beneficiaries of the actions, B and A ($c < b'$), and t and t' for the times of the actions.

part of the evolution of our species casts at least some doubt on the importance of reciprocal altruism for understanding the evolution of altruism.

A better reply to the charge that reciprocal altruism fails to explain a range of human altruistic behaviors consists in extending the notion of reciprocal altruism. Evolutionary biologist Richard Alexander (1987) has done precisely this with the notion of indirect reciprocity.[29] Roughly, the idea is that a trait that benefits another individual will be favored by natural selection if the possession of this trait increases the probability of benefiting from the altruism of a third party. One way to characterize indirect reciprocity is in terms of reputation. The possessor of the altruistic trait increases her reputation and people with a good reputation are the target of others' altruism (see Figure 1.2).

Prominent researchers, including Trivers himself, have proposed that while originally developed to explain altruism in a large range of species (including some species of fish), reciprocal altruism and indirect reciprocity also explain the evolution of morality in humans. Alexander puts it succinctly (1987: 77): "Moral systems are systems of indirect reciprocity."[30,31]

[29] The theory of indirect reciprocity has been developed by, among others, the mathematician Karl Sigmund and the theoretical biologist Martin Nowak, but a detailed discussion of their work is beyond the scope of this chapter (see, e.g., Nowak & Sigmund, 1998, 2005; Panchanathan & Boyd, 2004; for critical discussion, see Leimar & Hammerstein, 2001; Panchanathan & Boyd, 2003).

[30] Alexander's view is in fact more complex. For him, group selection is another cause of the evolution of morality.

[31] Joyce concurs, writing: "My own judgment is that . . . the process that most probably lies behind [the emergence of an innate faculty for making moral judgments] is indirect reciprocity, but it is not an objective of this book to advocate this hypothesis with any conviction" (2006: 44). Nowak & Sigmund (2005) make a similar claim, but their work is perhaps better interpreted as an instance of the second explanatory project that goes under the heading "the evolution of morality" (Section 2).

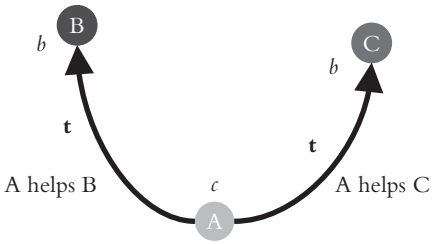

Figure 1.3. Public goods
Note: c stands for the cost of the altruistic action for the agent A, *b* for the benefit bestowed upon the beneficiaries of the actions, B and C, and *t* for the time of the action.

In spite of its impressive pedigree, the hypothesis that reciprocal altruism or indirect reciprocity selected for moral cognition, as a distinct kind of normative cognition, faces serious challenges. Following Sripada (2005), we highlight three shortcomings of this hypothesis. First, reciprocal altruism and indirect reciprocity are supposed to explain the evolution of behaviors in pairwise interactions, while moral behavior does not always take place in the context of pairwise interactions. For instance, morally sanctioned altruistic behaviors often benefit a large number of people, as is illustrated by the sacrifice of soldiers for their country.

One could propose to extend Trivers's reciprocal altruism to interactions that do not take place in pairs. This kind of situation is typically modeled by means of a public-goods game (aka, *n*-person prisoner's dilemma) (Figure 1.3).

Thus one could examine whether a trait that benefits the other members of the group conditional on the altruism of all the other members (or of a specific proportion of these group members) could be favored by natural selection, if interactions between the members of this group lasted long enough.[32] However, Boyd and Richerson (1988) have shown that natural selection is unlikely to favor this kind of trait (for critical discussion, see Johnson, Price, & Takezawa, 2008). Their model presents a dilemma. On the one hand, if altruists were to benefit others only when every other member of the group they belong to behaves altruistically, they would be unlikely to ever behave altruistically when this group is large and they would not reap the benefits of long-term cooperation. Their fitness would then not be higher than the fitness of non-altruists. On the other hand, if altruists were to behave altruistically when most (in contrast to all) members of their group behave altruistically,

[32] Altruistic behavior is often called "cooperation," and altruists are often called "cooperators" in the literature about the evolution of altruism.

they would then behave altruistically even when their group included some non-altruists. The fitness of altruists would then be lower than the fitness of non-altruists. Thus, on both horns of this dilemma, altruists have a lower fitness than non-altruists. Consequently, reciprocal altruism cannot explain the evolution of altruism in large groups.

Second, reciprocal altruism and indirect reciprocity are designed to explain the evolution of altruism (in a biological sense: acting in a way that increases others' fitness) while many moral norms have little to do with altruism because they do not regulate social interactions. For example, there are various apparently moral prohibitions against eating particular types of food. But it is very unclear how one can extend these two evolutionary mechanisms to account for the evolution of moral norms—like food taboos—that are not related to altruism.[33]

Third, neither reciprocal altruism nor indirect reciprocity can account for the link between morality and punishment. If reciprocal altruism or indirect reciprocity really explained the evolution of morality, people would be disposed to exclude from cooperation agents who violate moral norms. However, when an agent commits morally reprehensible actions, rather than merely terminating cooperation with this agent, people are disposed to punish her.[34] A large body of evidence in psychology and in behavioral economics highlights the link between punishment and morality. For instance, Haidt and Sabini (2000) showed subjects several videos of unjust actions. When asked to decide between various endings, subjects were more satisfied when the agent suffered for her action than when the victim pardoned the agent (see Fehr & Fischbacher, 2004 for consistent behavioral evidence).

To summarize, because the main adaptationist models of the evolution of morality appeal to direct or indirect reciprocity, they seem badly tailored to account for three key properties of moral cognition: moral norms do not exclusively (nor even primarily) bear on pairwise interactions; many moral norms have nothing to do with altruism; and violations of norms are punished.

[33] One might object that morality evolved for governing reciprocal interactions and, once present, came to govern other behaviors, such as incest avoidance. It is indeed common that a trait that was selected for a given function is put to other uses. Although we do not know any decisive reason to reject this hypothesis, it strikes us as an ad hoc extension of the hypothesis that reciprocal altruism or indirect reciprocity explains the evolution of morality.

[34] One could perhaps reject the distinction between punishing a norm violator and terminating cooperative interactions with a norm violator by arguing that the latter is a form of punishment. In reply, we highlight the importance of the distinction between punishing and terminating a cooperative relation. The former consists in imposing a cost on the norm violator, the latter in preventing the norm violator from gaining future benefits. Given that imposing a cost on the norm violator is typically costly (see Section 2), it is puzzling that people would go out of their way to punish others, while there is no mystery why people would stop cooperating when cooperating is not in their best interest.

For these three reasons, we view these models with skepticism. Their existence provides little support to the claim that moral cognition evolved.

3.3. *Distinctive Properties of Evolved Traits*

In addition to providing adaptationist models, it is also common to argue for the evolution of a trait by showing that this trait possesses some properties that are distinctive either of evolved traits or of adaptations. Here we focus successively on three alleged properties: moral norms are a cultural universal; complex moral norms are acquired in spite of impoverished stimuli; and very young children reason about norms and have harm-related emotions.

3.3.1. *Universality of Moral Norms* It is often said that there is no culture without a system of moral norms. Joyce writes (2006: 134): "[m]orality (by which I mean the tendency to make moral judgments) exists in all human societies we have ever heard of." Similarly, Hauser contends that "[t]he principles [of the moral faculty] constitute the universal moral grammar, a signature of the species" (2006: 53). And the universality of moral norms is taken to be evidence that it evolved.[35]

To evaluate this argument properly, it is important to keep in mind that the claim that morality is present in every culture is not just the assertion that one finds norms in every culture. Rather, it asserts that in every culture, one finds norms that possess the properties that distinguish moral norms from other kinds of norms according to the characterization used by a given evolutionary researcher. Thus, when Joyce asserts that morality is present in every culture, he is claiming that in every known culture, one finds norms that have most of the seven properties he uses to single out moral norms from other kinds of norms (see above).

But why should we believe that moral norms are present in all known cultures? Researchers often bluntly assert this claim (e.g. Dwyer, 2006: 237) or illustrate it with ancient or exotic codes of norms. Thus Joyce refers to the norms in the Egyptian Book of the Dead and in the Mesopotamian epic of Gilgamesh (2006: 134–135). We find this casual use of the anthropological and historical literature problematic. The problem is not so much that these researchers have not examined a large number of cultures and historical periods to substantiate the claim that moral norms are ancient and pancultural. Rather, the problem is that because they fail to clearly distinguish *norms* from *moral*

[35] For a brief discussion of the evidential connection between universality and evolution, see Section 2.3.

norms, the evidence they allude to merely supports the claim that norms are universal, but not the much more controversial claim that moral norms are universal (see Stich, 2008 for a similar point).

What anthropology and history show beyond reasonable doubt is that all known cultures have norms. In all cultures, some actions are prohibited while others are mandatory and some character traits are disvalued while others valued. Sanctions and rewards for behaviors and personal attributes are good candidates for being cultural universals. But because moral norms are conceived as a distinct type of norm and moral judgments as a distinct kind of normative judgment, the universality of norms should not be confused with the universality of *moral* norms. For a given researcher (e.g. Hauser, Dwyer or Joyce) to support the hypothesis that moral norms are universal thus requires far more than citing exotic and ancient codes. It requires him or her to show that such codes amount to an expression of morality, that is, to show that the norms in these codes possess the properties that distinguish moral norms from other kinds of norms according to the researcher's rich characterization of moral norms and judgments. However, to our knowledge, the relevant research has not been done. Furthermore, the richer the characterization of moral norms and judgments is, the less likely it is that norms in other cultures will count as moral norms and thus that moral norms will be a universal.

Let's illustrate this point with an example. In the sixth century, the Catholic Church prohibited Christians from being buried with their wealth and it recommended (but did not require) that part of one's wealth be given to the Church (Duby, 1996: 56–58). This is a clear example of a norm. But the existence of this norm in Europe fifteen centuries ago provides no support whatsoever for the hypothesis that moral norms are universal since it is unclear whether it is a moral norm—that is, since it is unclear whether it possesses the properties that distinguish moral norms. Instead of merely noting that this norm exists, Joyce and other researchers would have to show that it possesses the properties that (for them) distinguish moral cognition from other kinds of normative cognition. This is a very difficult task, and it is far from obvious whether the norms found in other cultures possess the properties that characterize moral norms according to their rich characterization.[36]

[36] Joyce has recently defended the claim that morality is universal (Joyce, 2008b). He asserts that there is probably no society where all the norms are prudential or hypothetical, noting that whenever norm violations are viewed as transgressions and as punishable, the norms violated are not merely hypothetical or prudential. He views this as evidence that moral norms are universal. We are not convinced by this argument, for one cannot infer that *moral* norms are universal from the fact that *categorical* norms are universal.

What might explain philosophers', psychologists', and anthropologists' confusion between the universality of norms and the universality of moral norms is the fact that they find in various cultures and times some norms that are somewhat similar to the norms they themselves view as moral. For instance, many cultures have norms against harming others, such as prohibition against in-group harm. But this fact does not show that all cultures have a system of moral norms, for, again, it is unclear whether these norms are moral norms in all the cultures in which they hold.

3.3.2. *The Moral/Conventional Distinction* The second piece of evidence adduced to support the hypothesis that morality evolved relies on the research on the moral/conventional distinction by developmental psychologist Elliot Turiel (Dwyer, 2006: 239–242; Joyce, 2006: 134–137; Hauser, 2006: 291).[37] In substance, Turiel and colleagues argue that very early on, and panculturally, children distinguish two types of norms, called "moral norms" and "conventional norms." Moral norms are those norms that are judged to hold independently from the authority of any individual or institution, that are judged to be universally applicable, that are justified by appeal to the harm done to others, to their rights, or to justice, and whose violations are judged to be serious. Conventional norms are those norms whose force depends on authority, that are judged to be only locally applicable, that are justified by reference to convention, and whose violations are judged to be less serious than the violations of moral norms.

Joyce concludes that "[t]hese results from developmental psychology strongly suggest that the tendency to make moral judgments is innate" (2006: 137). The argument from the universality and early development of the so-called moral/conventional distinction to the evolution of morality is best viewed as a poverty of the stimulus argument (Dwyer, 1999, 2006; Mikhail, 2000). According to this type of argument, developed most famously by Chomsky (1975), the fact that a trait, such as the capacity to speak a language, develops reliably, while the environmental stimuli are variable and impoverished, is evidence that this trait is innate (for discussion, see Cowie, 1999; Laurence & Margolis, 2001; Pullum & Scholz, 2002). Innateness is then often taken to be evidence that the trait under consideration is the product of evolution or sometimes that it is an adaptation.[38] It is tempting to apply this form of argument

[37] See, e.g., Turiel (1983); Nucci (2001); Smetana (1981); Blair (1995); Smetana et al. (1999).

[38] The evidential connection between being innate and having evolved (or being an adaptation) is not straightforward. The acquisition of some evolved traits involves learning, while some innate traits (such as some genetic diseases) are not evolved. For the sake of the argument, we take for granted here

to the distinction between moral norms and conventional norms. Turiel and others have argued that even very young children grasp the distinction between moral norms and other kinds of norms. It is dubious whether young children could have learned this distinction because the evidence needed to learn is often missing and because, when it is not missing, it is unreliable. For instance, Dwyer notes that caregivers' reactions to norms violations are unlikely to distinguish the two types of norms because, as she puts it (2006: 240), "(s)ome parents get just as hot under the collar about conventional transgressions as they do about moral transgressions." Furthermore, explicit moral instruction would not distinguish between different kinds of norms because moral norms are not linguistically distinguished from other norms: consider, e.g., "You ought to put your fork on the left of your plate" and "You ought to keep your promises." One might then conclude that the distinction between moral norms and conventional norms is innate.

The poverty of the stimulus argument about the so-called moral/conventional distinction is unsound. The research on this distinction has a long and respectable history, and it is received wisdom in much of contemporary psychology. Recently, however, a growing body of evidence has emerged that challenges the research tradition on the distinction between moral and conventional norms in psychology (Gabennesch, 1990; Haidt et al., 1993; Kelly et al., 2007).[39]

First, as Gabennesch (1990) has convincingly argued, the common wisdom—endorsed by Dwyer and others—that very early on, children view some norms as social conventions is poorly supported by the evidence. Carter and Patterson (1982) found that half of their second- and fourth-grader subjects judged that table manners (e.g. eating with one's fingers) were not variable across cultures and that they were authority-independent. Similarly, Shweder and colleagues (1987: 35) concluded that among American children under 10, "there [was] not a single practice in [their] study that is viewed predominantly in conventional terms" (see Gabennesch, 1990 for many other references). Because many children do not understand early on that some norms are conventional, the poverty of the stimulus argument about the moral/conventional distinction mischaracterizes the nature of the moral knowledge of young children. Furthermore, because people come to understand slowly and rather late

that this connection can be drawn. In addition, we bracket the debate about what innateness is and about whether the notion of innateness is confused (for discussion, see, e.g., Samuels, 2002; Mallon & Weinberg, 2006; Griffiths, Machery, & Linquist, 2009).

[39] For further critical discussion of the poverty of the stimulus argument under consideration, see Nichols (2005) and Prinz (2008). We also note that we do not reject poverty of the stimulus arguments in general.

in their adolescence that some norms are mere social conventions, the amount of evidence children and adolescents might rely on to come to understand this distinction (whatever it amounts to—see below) is much less impoverished than is assumed by the poverty of the stimulus argument.

More important, the research on the so-called moral/conventional distinction assumes that being authority-independent, being universal, being associated with serious violations, and being justified by appeal to harm, justice, or rights form a cluster of co-occurring properties (Kelly et al., 2007). That is, it is assumed that when a norm is judged to be authority-independent, it is also judged to be universal, it is justified by appeal to harm, justice, or rights, and the violations of this norm are judged to be serious. By contrast, when a norm is judged to be authority-dependent, it is not judged to be generalizable to other cultures, it is justified by appeal to conventions, and the violations of this norm are judged to be less serious. However, research shows that this assumption is unsubstantiated. People judge some actions (e.g. having sex with a dead chicken before eating it or cleaning the bathroom with the national flag) to be serious and authority-independent, but do not justify the relevant norms by appeal to harm, justice, or rights (Haidt et al., 1993). People are also sometimes reluctant to generalize to other cultures norms that are justified by appeal to harms, and they sometimes view these norms as authority-dependent (Kelly et al., 2007; but see Sousa, 2009; Sousa, Holbrook, & Piazza, 2009 for discussion). It thus appears that the four properties assumed to set apart moral norms may not form a cluster of co-occurring properties at all.[40] If this is the case, it is unclear what the claim that early on children distinguish moral norms from conventional norms amounts to, putting into jeopardy the poverty of the stimulus argument about the moral/conventional distinction.

3.3.3. *Developmental Evidence* In addition to the two alleged pieces of evidence discussed above (the presence of moral norms in every culture and the early development of the moral/conventional distinction), other aspects of human psychological development have been mentioned as evidence for the evolution of morality. However, as we now argue, they do not constitute evidence that moral cognition, understood again as a specific kind of normative cognition, evolved, rather than evidence that normative cognition evolved.

We have seen in Section 2 that children understand deontic conditionals, such as ''It's not safe outside for the squeaky mouse, so all squeaky mice *must*

[40] One could object that although the four properties assumed to set apart moral norms do not necessarily occur together, they still *tend* to co-occur, forming something like a homeostatic cluster (Boyd, 1991). However, Kelly and colleagues' work tentatively suggests that these properties do not tend to co-occur, since most possible combinations seem to occur.

stay in the house'' much earlier and much better than indicative conditionals, such as ''It's not safe outside for the squeaky mouse, so all squeaky mice *are* in the house'' (Cummins, 1996a; Harris & Núñez, 1996). One might be tempted to argue that the capacity to identify the violation of deontic conditionals early on provides evidence for the evolution of morality (Joyce, 2006; Dwyer, 2006).

This is certainly a very interesting finding, one that suggests that normative cognition might be the product of evolution by natural selection (as noted in Section 2). However, it is unclear how this finding is supposed to support the idea that moral cognition proper, understood as a specific kind of normative cognition, rather than normative cognition in general, evolved, since the norms in the stories presented to children by Cummins and by Harris and Núñez were not moral. More generally, deontic conditionals are not exclusively used in specifically moral reasoning.

One could perhaps argue that infants' empathic reaction to others' suffering and the early development of helping behaviors in children provide evidence for the evolution of morality. For instance, Dwyer concludes that ''this work strongly suggests that some basic moral capacities are in place quite early in development'' (2006: 237). However, again, it is unclear how the early development of empathy and of helping behaviors is supposed to support the hypothesis that morality is a product of evolution. Certainly, empathy is morally sanctioned in modern, Western cultures and helping is often morally prescribed. But all this shows is that empathy and some behavioral tendencies that are morally sanctioned in modern, Western cultures are present at an early stage of children's psychological and behavioral development. This is perfectly consistent with moral norms being a culture-specific kind of norms and with moral cognition being a culture-specific kind of normative cognition that recruits early developing, maybe evolved psychological traits, such as empathy and some behavioral tendencies.

To summarize, while many philosophers, psychologists, and anthropologists have claimed that morality is a product of the evolution of the human species, the evidence for this claim is weak at best. First, we do not know whether moral norms are present in every culture: because researchers endorse rich characterizations of what moral norms are, it is not obvious that norms that have the distinctive properties of moral norms will be found in every culture, and, in any case, researchers have simply not shown that, in numerous cultures, there are norms that fit some rich characterization of moral norms. Second, the claim that early on children display some complex moral knowledge in spite of variable and impoverished environmental stimuli is based on the research on the moral/conventional distinction. Although this research remains widely

accepted in much of psychology, a growing body of evidence has highlighted its shortcomings. Third, the other pieces of evidence often cited in the literature on the evolution of morality do not constitute evidence that moral norms and moral judgments, understood as a specific type of norms and normative judgments, evolved, rather than evidence that normative cognition evolved.

3.4. *Summary: The Evolution of a Specifically Moral Normativity*

In this section, we have focused on the idea that moral cognition, conceived as a distinct kind of normative cognition, evolved. We have argued that the scenarios typically advanced to explain the selection of this specific form of normative cognition are unconvincing, and that the arguments and evidence commonly adduced to support the hypothesis that morality evolved are at best inconclusive.

This conclusion is philosophically significant. As noted already, it is commonly argued that the evolution of morality undermines the authority of moral norms. At the end of Section 2, we noted that this argument cannot plausibly hang on the hypothesis that normative cognition in general evolved (the second interpretation of the claim that morality evolved) although this hypothesis is supported by a small, but suggestive, body of evidence. Rather, it should hang on the third interpretation that morality evolved—i.e. moral cognition, conceived as a distinct kind of normative cognition, is the product of evolution—and this is indeed the way it has typically been presented.

Most philosophers have focused on evaluating the truth of the conditional, "If morality (understood as a particular type of normative cognition) evolved, the authority of moral norms is undermined" (e.g. Joyce, 2006; Street, 2006, 2008; Copp, 2008). However, no agreement has been reached on whether this conditional should be accepted. Our discussion shows that this lack of agreement might not matter since it turns out that there is little reason to believe that moral cognition, understood as a particular type of normative cognition, evolved. The claim that the authority of moral norms is undermined by the evolution of morality therefore depends on an unsupported premise, and so it does not threaten the authority of moral norms.

4. Conclusion

So, did morality evolve? We have shown that this question has no single answer, because it is understood in various ways. Some researchers focus on the evolutionary history of specific components of moral psychology. So

understood, it is uncontroversial that morality evolved: although establishing that some particular morally relevant trait has evolved can be especially difficult, there is little doubt that numerous traits have a long evolutionary history. However, philosophers are unlikely to be moved by this conclusion, since it is unclear whether it has any philosophically significant implication.

By arguing that morality evolved, other researchers contend that normative cognition is an adaptation. We have argued that, although somewhat speculative, this claim is supported by a small, but suggestive, body of sociological and psychological evidence as well as by a robust set of how-possible evolutionary models. We view the fact that norms are ancient and universal and that from a very early age on, people are distinctively adept at reasoning about normative matters, as evidence that normative cognition evolved by natural selection. But again, we argued that this conclusion is cold comfort to philosophers hoping to draw conclusions undermining the authority of moral norms from the evolution of morality.

Finally, other researchers characterize moral cognition as a distinct kind of normative cognition, which includes the grasp of a specific kind of norms (i.e. moral norms) and a capacity to make a specific kind of normative judgments (i.e. moral judgments). They then endorse the provocative claim that we evolved to grasp this specific kind of norms and to make this specific kind of normative judgments. By contrast, we have argued that the evidence usually adduced to support the hypothesis that morality (so characterized) evolved is far from conclusive. While some adaptationist models, inspired by evolutionary biologists' research on altruism, are touted as suggesting that morality is an adaptation, we have argued that they are not well tailored to explain how morality evolved. Researchers also assert that the universality and innateness of morality show that it evolved. But a critical look at the evidence reveals that it is unclear whether morality is universal and innate.

As we noted in the introduction, the hypothesis that morality evolved is often assumed to have significant philosophical implications. The evolution of morality features in arguments attempting to justify specific norms and in various skeptical arguments about morality. Although our discussion has not directly focused on evaluating these arguments, it has led us to skepticism about a crucial premise. While the first reading of the claim that morality evolved is uncontroversial, and while its second reading is supported by a small, but suggestive, body of evidence, these two readings do not seem to yield significant philosophical payoffs. While the third reading—morality, understood as a distinct type of normative cognition, evolved—is more likely to yield such payoffs, on close consideration, it turns out to be empirically unsupported.

References

Alcock, J. (2001). *The Triumph of Sociobiology*. New York: Oxford University Press.

Alexander R. D. (1987). *The Biology of Moral Systems*. Hawthorne, NY: Aldine de Gruyter.

Allen, C., & Bekoff, M. (2005). Animal play and the evolution of morality: An ethological approach. *Topoi, 24*, 125–135.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics, 116*, 293–312.

Bekoff, M. (2001). Social play behavior: Cooperation, fairness, trust, and the evolution of morality. *Journal of Consciousness Studies, 8*, 81–90.

——(2004). Wild justice and fair play: Cooperation, forgiveness, and morality in animals. *Biology and Philosophy, 19*, 489–520.

Benedict, R. (1946). *The Chrysanthemum and the Sword: Patterns of Japanese Culture*. Boston, MA: Houghton Mifflin.

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*, 1–29.

Boehm, C. (1982). The evolutionary development of morality as an effect of dominance behavior and conflict interference. *Journal of Social and Biological Structures, 5*, 413–421.

——(1999). *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior, 10*, 95–121.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies, 61*, 127–148.

Boyd, R., & Mathew, S. (2007). A narrow road to cooperation. *Science, 316*, 1858–1859.

Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology, 132*, 337–356.

——(1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*, 171–195.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences, 100*, 3531–3535.

Bowles, S., & Gintis, H. (1998). The moral economy of community: Structured populations and the evolution of prosocial norms. *Evolution and Human Behavior, 19*, 3–25.

Brandon, R. (1990). *Organism and Environment*. Princeton, NJ: Princeton University Press.

Bräuer, J., Call, J., & Tomasello, M. (2006). Are apes really inequity averse? *Proceedings of the Royal Society of London B, 273*, 3123–3128.

Brigandt, I. (2003). Homology in comparative, molecular, and evolutionary developmental biology: The radiation of a concept. *Journal of Experimental Zoology (Molecular and Developmental Evolution), 299*, 9–17.

Brosnan, S. F. (2006). Nonhuman species' reactions to inequity and their implications for fairness. *Social Justice Research, 19*, 153–185.

Brosnan, S. F., & de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature, 13*, 129–152.

——(2003). Monkeys reject unequal pay. *Nature, 425*, 297–299.

——(2004). Reply to Henrich and Wynne. *Nature, 428*, 140.

Brosnan, S. F., Freeman, C., & de Waal, F. B. M. (2006). Partner's behavior, not reward distribution, determines success in an unequal cooperative task in capuchin monkeys. *American Journal of Primatology, 68*, 713–724.

Brosnan, S. F., Schiff, H. C., & de Waal, F. B. M. (2005). Tolerance for inequity may increase with closeness in chimpanzees. *Proceedings of the Royal Society of London, Series B, 1560*, 253–258.

Brown, D. E. (1991). *Human Universals*. New York: McGraw-Hill.

Carter, D. B., & Patterson, C. J. (1982). Sex roles as social conventions: The development of children's conceptions of sex-role stereotypes. *Child Development, 18*, 812–824.

Casebeer, W. D. (2003a). *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*. Cambridge, MA: MIT Press.

——(2003b). An argument for "new wave" Aristotelianism. *Politics and the Life Sciences, 22*, 67–69.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.

Copp, D. (2008). Darwinian skepticism about moral realism. *Philosophical Issues, Interdisciplinary Core Philosophy, 18*, 186–206.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187–276.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (ed.), *The Handbook of Evolutionary Psychology*. Hoboken, NJ: Wiley, 584–627.

——(2008a). Can a general deontic logic capture the facts of human moral reasoning? How the mind interprets social exchange rules and detects cheaters. In W. S. A. Armstrong (ed.), *Moral Psychology, volume 1: The Evolution of Morality*. Cambridge, MA: MIT Press, 53–120.

——(2008b). Can evolutionary psychology assist logicians? A reply to Mallon. In W. S. A. Armstrong (ed.), *Moral Psychology, volume 1: The Evolution of Morality*. Cambridge, MA: MIT Press, 131–136.

——(2008c). When falsification strikes: A reply to Fodor. In W. S. A. Armstrong (ed.), *Moral Psychology, volume 1: The Evolution of Morality*. Cambridge, MA: MIT Press, 143–164.

Cowie, F. (1999). *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.

Cronk, L. (1994). Evolutionary theories of morality and the manipulative use of signals. *Zygon: Journal of Religion and Science, 29*, 81–101.

Cummins, D. D. (1996a). Evidence of deontic reasoning in 3- and 4-year-olds. *Memory and Cognition, 24*, 823–829.

Cummins, D. D. (1996b). Evidence for the innateness of deontic reasoning. *Mind & Language, 11*, 160–190.

D'Arms, J. (2000). When evolutionary game theory explains morality, what does it explain? *The Journal of Consciousness Studies, 7*, 296–300.

——(ms). Self-righteous anger: A case study in evolutionary ethics.

Darwall, S., Gibbard, A., & Railton, P. (1992) Toward fin de siècle ethics: Some trends. *Philosophical Review, 101*, 115–189.

Darwin C. (1871). *The Descent of Man and Selection in Relation to Race*. 1874 edn. London: John Murray.

Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.

Draper, P., & Belsky, J. (1990). Personality development in evolutionary perspective. *Journal of Personality, 58*, 141–162.

Dubreuil, D., Gentile, M. S., & Visalberghi, E. (2006). Are capuchin monkeys (Cebus apella) inequity averse? *Proceedings of the Royal Society of London B, 273*, 1223–1228.

Duby, G. (1996). *Féodalité*. Paris: Gallimard.

Dunbar, R. I. M. (1996). *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Dwyer, S. (1999). Moral competence. In K. Murasugi & R. Stainton (eds.), *Philosophy and Linguistics*. Boulder, CO: Westview Press, 169–190.

——(2006). How good is the linguistic analogy? In P. Carruthers, S. Laurence, & S. Stich. *The Innate Mind: Culture and Cognition*. Oxford: Oxford University Press, 237–256.

Farber, P. (1994). *The Temptation of Evolutionary Ethics*. Berkeley, CA: University of California Press.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63–87.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140.

Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? In P. Hammerstein (ed.), *Genetic and Cultural Evolution of Cooperation*. Cambridge, MA: MIT Press, 55–82.

Fessler, D. M. T. (1999). Toward an understanding of the universality of second order emotions. In A. Hinton (ed.), *Beyond Nature or Culture: Biocultural Approaches to the Emotions*. New York: Cambridge University Press, 75–116.

——(2004) Shame in two cultures: Implications for evolutionary approaches. *Journal of Cognition and Culture, 4*, 207–262.

——(2007) From appeasement to conformity: Evolutionary and cultural perspectives on shame, competition, and cooperation. In J. L. Tracy, R. W. Robins, & J. P. Tangney (eds.), *The Self-conscious Emotions: Theory and Research*. New York: Guilford Press, 174–193.

Fessler, D. M. T., & Haley, K. J. (2003) The strategy of affect: Emotions in human cooperation. In P. Hammerstein (ed.), *The Genetic and Cultural Evolution of Cooperation*. Cambridge, MA: MIT Press, 7–36.

Fiske, A. P. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York: Free Press.

Flack, J. C., & de Waal, F. B. M. (2000). "Any animal whatever": Darwinian building blocks of morality in monkeys and apes. *Journal of Consciousness Studies, 7*, 1–29.

Fodor, J. A. (2000). Why we are so good at catching cheaters. *Cognition*, *75*, 29–32.

——(2008). Comment on Cosmides and Tooby. In W. S. A. Armstrong (ed.), *Moral Psychology, volume 1: The Evolution of Morality*. Cambridge, MA: MIT Press, 137–142.

Foster, M. W., Gilby, I. C., Murray, C. M., Johnson, A., Wroblewski, E. E., & Pusey, A. E. (2009). Alpha male chimpanzee grooming patterns: Implications for dominance "style". *American Journal of Primatology*, *71*, 136–144.

Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. New York: W. W. Norton & Company.

Gabennesch, H. (1990). The perception of social conventionality by children and adults. *Child Development, 61*, 2047–2059.

Gächter, S., & Falk, A. (2002). Reputation or reciprocity? Consequences for labour relations. *Scandinavian Journal of Economics, 104*, 1–25.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans, *Evolution and Human Behavior, 24*, 153–172.

Gould, S. J., & Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society B, 205*, 581–598.

Griffiths, P. E. (2006). Function, homology, and character individuation. *Philosophy of Science, 73*, 1–25.

Griffiths, P. E., Machery, E., & Linquist, S. (2009). The vernacular concept of innateness. *Mind & Language, 24*, 605–630.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (eds.), *Handbook of Affective Sciences*. Oxford: Oxford University Press, 852–870.

Haidt, J., & Sabini, J. (2000). What exactly makes revenge sweet? Unpublished manuscript.

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–628.

Hamilton, W. (1964). The evolution of social behavior. *Journal Theoretical Biology, 7*, 1–52.

Harris, P. L., & Núñez. M. (1996). Understanding of permission rules by preschool children. *Child Development, 67*, 1572–1591.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A., & Sigmund, K. (2007). Via freedom to coercion: The emergence of costly punishment. *Science, 316*, 1905–1907.

Hauser, M. D. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Ecco.

Hauser, M. D., Cushman, F., Young, L., Jin, K.-X. R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22*, 1–21.

Henrich, J. (2004). Inequity aversion in capuchins. *Nature, 428*, 139.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208*, 79–89.

Henrich, H., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2004). *Foundations of Human Sociality*. New York: Oxford University Press

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J. Q., Smith, N., & Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences, 28*, 795–855.

Hill, K. (2003). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature, 13*, 105–128.

Huxley, T. H. (1894/1989). *Evolution and Ethics*. J. Paradis and G. C. Williams (eds.). Princeton, NJ: Princeton University Press.

Irons, W. (1991). How did morality evolve? *Zygon, 26*, 49–89.

Johnson, D. D. P., Price, M. E., & Takezawa, M. (2008). Renaissance of the individual: Reciprocity, positive assortment, and the puzzle of human cooperation. In C. Crawford & D. Krebs (eds.), *Foundations of evolutionary psychology*. New York: Lawrence Erlbaum, 331–352.

Johnson, D. D. P., Stopka, P., & Knights, S. (2003). The puzzle of human cooperation. *Nature, 421*, 911–912.

Joyce, R. (2000). Darwinian ethics and error. *Biology and Philosophy, 15*, 713–732.

——(2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

——(2008a). Précis of *The Evolution of Morality*. *Philosophy and Phenomenological Research, 77*, 213–218.

——(2008b). Replies. *Philosophy and Phenomenological Research, 77*, 245–267.

Kelly, D., Stich, S. P., Haley, K. J., Eng, S., J., & Fessler, D. M. T. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language, 22*, 117–131.

Kitcher, P. (1985). *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge, MA: MIT Press.

——(1998). Psychological altruism, evolutionary origins, and moral rules. *Philosophical Studies, 98*, 283–216.

——(2006a). Biology and ethics. In D. Copp (ed.), *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press, 163–185.

——(2006b). Between fragile altruism and morality: Evolution and the emergence of normative guidance. In G. Boniolo & G. de Anna (eds.), *Evolutionary Ethics and Contemporary Biology*. New York: Cambridge University Press, 159–177.

Krebs, D. L. (2005). The evolution of morality. In D. Buss (ed.), *The Handbook of Evolutionary Psychology*. Hoboken, NJ: John Wiley & Sons, 747–771.

Kropotkin, P. (1902). *Mutual Aid: A Factor in Evolution*. London: McClure Phillips and Co.

Lahti, D. C. (2003). Parting with illusions in evolutionary ethics. *Biology and Philosophy, 18*, 639–651.

Laurence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal for Philosophy of Science, 52*, 217–276.

Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London (B), 268*, 745–753.

Levy, N. (2004). *What Makes us Moral? Crossing the Boundaries of Biology*. Oxford: Oneworld Publications.

——(2007). The responsibility of the psychopath revisited. *Philosophy, Psychiatry, & Psychology, 14*, 129–138.

Lloyd, E. A. (2005). *The Case of the Female Orgasm: Bias in the Science of Evolution*. Cambridge, MA: Harvard University Press.

Machery, E. (Forthcoming). Discovery and confirmation in evolutionary psychology. In J. J. Prinz (ed.), *Oxford Handbook of Philosophy of Psychology*. Oxford: Oxford University Press.

Mallon, R. (2008). Ought we to abandon a domain-general treatment of "ought"? In W. S. A. Armstrong (ed.), *Moral Psychology, volume 1: The Evolution of Morality*. Cambridge, MA: MIT Press, 121–130.

Mallon, R., & Weinberg, J. (2006). Innateness as closed-process invariantism. *Philosophy of Science, 73*, 323–344.

McAdams, R. H. (1997). The origin, development, and regulation of social norms. *Michigan Law Review, 96*, 338–443.

McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: A new interpretation of the origin of modern human behavior. *Journal of Human Evolution, 39*, 453–563.

Mikhail, J. (2000). Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in *A theory of justice*. PhD thesis, Cornell University, p. 375.

Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.

——(2005). Innateness and moral psychology. In P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind: Structure and Content*. New York: Oxford University Press, 353–370.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*, 573–577.

——(2005). Evolution of indirect reciprocity. *Nature, 437*, 1291–1298.

Nucci, L. (2001). *Education in the Moral Domain*. Cambridge: Cambridge University Press.

Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: The importance of standing for the evolution of indirect reciprocity. *Journal of Theoretical Biology, 224*, 115–126.

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432*, 499–502.

Preston, S. D., & de Waal, F. B. M. (2002). The communication of emotions and the possibility of empathy in animals. In S. Post, L. G. Underwood, J. P. Schloss & W. B. Hurlburt (eds.), *Altruistic Love: Science, Philosophy, and Religion in Dialogue*. New York: Oxford University Press, 284–308.

Prinz, J. J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.

——(2008). Acquired moral truths. *Philosophy and Phenomenological Research, 77*, 219–227.

Pullum, G. K., and Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review, 19*, 9–50.

Railton, P. (1999). Normative force and normative freedom: Hume and Kant, but no Hume *versus* Kant. *Ratio, 12*, 320–353.

Rangea, F., Horna, L., Viranyi, Z., & Hubera, L. (2009). The absence of reward induces inequity aversion in dogs. *Proceedings of the National Academy of Sciences, 106*, 340–345.

Richards, R. J. (1986). A defence of evolutionary ethics. *Biology and Philosophy, 1*, 265–293.

——(1989). Dutch objections to evolutionary ethics. *Biology and Philosophy, 4*, 331–343.

Richerson, P. J., & Boyd, R. (1998). The evolution of human ultra-sociality. In I. Eibl-Eibisfeldt & F. Salter (eds.), *Ideology, Warfare, and Indoctrinability*. Oxford: Berghan Books, 71–95.

——(1999). The evolutionary dynamics of a crude super organism. *Human Nature, 10*, 253–289.

——(2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Richerson, P. J., Boyd, R., & Henrich, J. (2003). The cultural evolution of human cooperation. In P. Hammerstein (ed.), *The Genetic and Cultural Evolution of Cooperation*. Cambridge MA: MIT Press, 357–388.

Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy." *Philosophical Psychology, 16*, 51–66.

Rottschaefer, W. A. (1991). The insufficiency of supervenient explanations of moral actions: Really taking Darwin and the naturalistic fallacy seriously. *Biology and Philosophy, 6*, 439–445.

——(1998). *The Biology and Psychology of Moral Agency*. Cambridge: Cambridge University Press.

Rottschaefer, W. A., & Martinsen, D. (1990). Really taking Darwin seriously: An alternative to Michael Ruse's Darwinian metaethics. *Biology and Philosophy, 5*, 149–173.

Ruse, M. (1986). *Taking Darwin Seriously*. Oxford: Basil Blackwell.

Ruse, M., & Wilson, E. O. (1985). Moral philosophy as applied science. In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 421–438.

Samuels, R. (2002). Nativism in cognitive science. *Mind & Language, 17*, 233–265.

Shweder, R. A., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan and S. Lamb (eds.), *The Emergence of Morality in Young Children*. Chicago, IL: University of Chicago Press, 1–82.

Singer, P. (1981). *The Expanding Circle: Ethics and Sociobiology*. Oxford: Oxford University Press.

——(2000). *A Darwinian Left: Politics, Evolution, and Cooperation*. New Haven, CT: Yale University Press.

Smetana, J. (1981). Preschool children's conceptions of moral and social rules. *Child Development, 52*, 1333–1336.

Smetana, J., Toth, S., Cicchetti, D., Bruce, J., Kane P., & Daddis, C. (1999). Maltreated and nonmaltreated preschoolers' conceptions of hypothetical and actual moral transgressions. *Developmental Psychology, 35*, 269–281.

Smith, E. A. (2004). Why do good hunters have higher reproductive success? *Human Nature, 15*, 343–364.

Sober, E. (1994). Prospects for an evolutionary ethics. In E. Sober, *From a Biological Point of View*. Cambridge: Cambridge University Press, 93–113.

Solnick, S. J.(2001). Gender differences in the ultimatum game. *Economic Inquiry, 39*, 189–200.

Sousa, P. (Forthcoming). On testing the "moral law." *Mind & Language*.

Sousa, P., Holbrook, C., & Piazza, J. (Forthcoming). The morality of harm. *Cognition*.

Spencer, H. (1892). *The Principles of Ethics*. London: Williams and Northgate.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 52,* 3–39.

Sripada, C. (2005). Punishment and the strategic structure of moral systems. *Biology and Philosophy, 20*, 707–789.

Sripada, C., & Stich, S. (2006). A Framework for the Psychology of Norms. In P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind: Culture and Cognition*. Oxford: Oxford University Press, 280–301.

Stich, S. P. (2008). Some questions about the evolution of morality. *Philosophy and Phenomenological Research, 77*: 228–236.

Stone, V. E. (2006). Theory of mind and evolution of social intelligence. In J. T. Cacioppo, P. S. Visser, & C. L. Pickett (eds.), *Social Neuroscience: People Thinking about Thinking People*. Cambridge, MA: MIT Press, 103–130.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 127*, 109–166.

——(2008). Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Studies, Interdisciplinary Core Philosophy, 18*, 207–228.

Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences, 99*, 11537–11542.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57.

Turiel, E. (1983). *The Development of Social Knowledge*. Cambridge: Cambridge University Press.

de Waal, F. B. M. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.

Waddington, C. H. (1942) *Science and ethics*. London: Allen & Unwin.

Williams, G. C. (1988). Huxley's evolution and ethics in sociobiological perspective. *Zygon, 23*, 383–407.

Wilson, D. S. (2002). *Darwin's Cathedral: Evolution, Religion and the Nature of Society*. Chicago, IL: University of Chicago Press.

van Wolkenten, M., Brosnan, S. F., & de Waal, F. B. M. (2007). Inequity responses of monkeys modified by effort. *Proceedings of the National Academy of Science*, *104*(47): 18854–18859.

Woolcock, P. G. (2000). Objectivity and illusion in evolutionary ethics: Comments on Waller. *Biology and Philosophy, 15*, 39–60.

Wright, R. (1994). *The Moral Animal*. New York: Pantheon Books.