Ethics and evolution
Richard Joyce

1. Evolutionary ethics

The field known as *evolutionary ethics* has burgeoned over the last couple of decades, yet its remit and defining concerns remain obscure to most. In terms of both conceptual clarification and labor, evolutionary ethics can be usefully divided into an empirical and a philosophical program. I will provide a quick overview of both, then return to each in turn in more detail.

Empirical evolutionary ethics:

Humans are (perhaps uniquely) moral creatures. In this context, this doesn't mean that humans are morally praiseworthy or admirable (or, for that matter, blameworthy or iniquitous); nor does it mean that humans are proper subjects of moral concern. Rather, it means that humans make moral judgments: We classify our world in terms of moral values, our actions in terms of moral rules, our character traits in terms of moral virtues and vices, and so forth. Where does this way of thinking come from? One may pose this question synchronically, and it is the job of moral psychology to answer that question—to reveal what faculties are involved in moral judgment. But one may also ask the question diachronically, in which case one investigates by what processes humans came to have those faculties involved in moral judgment in the first place. One possibility is that moral judgment is a relatively recent cultural invention that exploits various psychological faculties that evolved for other purposes. Another possibility is that there exist in the human mind mechanisms that evolved specifically to make moral judgment possible. On the latter hypothesis—which can be called "moral nativism"—moral judgment is a biological adaptation that emerged because this way of thinking provided our ancestors with some sort of reproductive advantage over their competition.

Moral nativism is an empirical matter. Of course, there may be concepts that require clarification prior to embarking on the investigation (most obviously, what is a *moral judgment*), so a certain amount of a priori examination is involved in the discussion. But, ideally, once these clarifications *au fauteuil* have been made, then the matter can be turned over to scientists. Part of what makes empirical evolutionary ethics exciting is the wide range of scientists whose skills are relevant to the hypothesis. Key contributions can be made by social and developmental psychologists, experimental economists, neuroscientists, geneticists, primatologists, anthropologists, comparative ethologists, and evolutionary biologists. A terminological oddity of the field is that though explicit discussion of moral nativism from any such scientist can be placed under the rubric "evolutionary ethics," this empirical enterprise is not a field of *ethics* in the traditional sense. Thus, an evolutionary psychologist, say, who advocates moral nativism, will often be taken to count as an

evolutionary ethicist but not as an ethicist. (Proponents of moral nativism include Alexander 1987; Krebs 2005; Dwyer 2006; Hauser 2006; Joyce 2006; Mikhail 2011; Kitcher 2011.)

Philosophical evolutionary ethics:

By contrast, philosophical evolutionary ethics *is* a subfield of ethics. The philosophical evolutionary ethicist proposes that facts about human evolution can help address certain perennial problems in moral philosophy. Moral nativism (if true) represents a prominent example of the kind of "fact about human evolution" that may have some bearing on moral philosophy. But even if moral nativism is rejected, there are other possibilities for philosophical evolutionary ethics. The uncontroversial fact that humans have evolved by natural selection to be social creatures, for example, might be appealed to in order to support or undermine certain ethical theses.

Some have given in to the temptation to call philosophical evolutionary ethics "*normative* evolutionary ethics," but this is misleading and it is instructive to note why. Moral philosophy is a wide-ranging and varied field, the outputs of only some areas of which should be thought of as *normative*. At one extreme lies applied ethics, which may offer definite practical advice on how to act in concrete scenarios (e.g., regarding a specific case of euthanasia). Normative ethics is the enterprise of building a general theory of moral action that is applicable across all or a large range of cases (e.g., utilitarianism). At the other end of the spectrum lies metaethics, which is concerned with a number of interrelated theoretical matters, such as the ontology of moral properties, the nature of moral language, and the epistemological status of moral judgments. Philosophical evolutionary ethics is the investigation of whether facts about human evolution might have some non-trivial constructive input to moral philosophy at any point on this vague spectrum. But this need not be a *normative* impact. Consider, for example, the longstanding metaethical debate over whether moral language functions assertorically or whether it performs some other speech act. Were some finding about the nature of human evolution to help settle this dispute—demonstrating, say, that moral utterances are assertions—this would hardly be a *normative* result.

This is important to note because philosophical evolutionary ethics is sometimes rejected out of hand on the grounds that it must at some point purport to derive "ought"-claims from descriptive premises—a move that allegedly commits a logical fallacy. But this is not so. For example, the aforementioned metaethical thesis—"Moral judgments are assertions"—is not an "ought"-claim. A claim about how the word "ought" functions in everyday speech is not itself an "ought"-claim. Moreover, even when the philosophical evolutionary ethicist does maintain that evolutionary considerations lend support to some normative claim or other, it is unlikely that he or she thinks that the descriptive (evolutionary) premises *alone* imply that result. And there is nothing fishy about drawing an "ought"-conclusion from premises some of which are descriptive and some of which also contain "ought"-statements. In short, the common accusation that philosophical evolutionary ethics must fail, because its advocates must be endeavoring to derive an "ought" from an "is," usually misidentifies what the philosophical evolutionary ethicist is attempting.

Philosophical evolutionary ethics can be divided into two opposed programs: vindicating and debunking. There are different kinds of vindication possible. One might seek to use evolutionary data to establish a positive result in applied ethics (e.g., "An act of euthanasia in such-and-such circumstances would be wrong"), or a result in normative ethics (e.g., "Utilitarianism is the best moral theory"), or in metaethics (e.g., "Moral properties objectively exist"). But not any positive metaethical result will count as a vindication, because some metaethical theses are considered to be, by their very nature, antithetical to morality. For example, the error theoretic view that moral judgments aim at the truth but systematically fail to secure it, is for obvious reasons considered as hostile to morality as a practice. (See Mackie 1977.) Thus if one could use evolutionary data to help establish the error theory, then although this would in some sense "vindicate" a certain metaethical view, it would be more natural to consider it as a form of evolutionary *debunking*. (The same could be said of the epistemological thesis that all moral judgments are unjustified.) In some circumstances it is difficult to say whether it is a vindication or debunking that is under discussion, for it depends on the viewer's perspective. Sometimes an evolutionary ethicist may aim to destroy one ethical viewpoint in order to replace it with another. Edward O. Wilson, for example, in collaboration with Michael Ruse, writes in debunking spirit that Darwinism shows us that "ethics as we understand it is an illusion fobbed off on us by our genes to get us to cooperate" (Ruse & Wilson 1985: 52); and yet elsewhere Wilson opines that a proper evolutionary understanding of ethics will "make possible the selection of a more deeply understood and enduring code of moral values" (Wilson 1978: 196).

A sketch of the bipartite field of evolutionary ethics having now been provided, the rest of this chapter basically repeats this pattern but in more detail. The treatment will be asymmetric, expending more than three times the ink on the former task than the latter. No particular view will be argued for; my aspirations do not extend beyond providing an overview (though, inevitably, a somewhat biased one) of the dialectic.

2. Moral nativism

One should not really speak of "moral nativism" in the singular, for there are numerous hypotheses deserving of the title. While moral nativism is usually identified with the thesis that human morality is innate, there are two glaring sources of indeterminacy: (i) What does "innate" mean in this context? (ii) Which trait(s) does "human morality" denote?

What is innateness?

Several commentators have noticed that different ideas are uncomfortably gathered under the term "innate" (Griffiths 2002; Mameli & Bateson 2007). One idea privileges the notion of a trait whose developmental emergence is relatively insensitive to environmental variation. Another idea focuses on a trait's being essential to species membership. Another important use of "innate" denotes a trait that has been selected for by natural selection. All three uses would need further tightening before they could be treated as scientifically respectable, but even these loose versions suffice to reveal non-equivalency. Cats' having four legs is both

developmentally robust and a feline adaptation, but is not an essential feature of being a cat. The genetic impairment Down's Syndrome is developmentally robust but is not an adaptation. Singing behavior is an adaptation for certain birds but sometimes requires crucial environmental input. And so forth.

When it comes to debates over moral nativism, different writers often use different conceptions of "innate," and even when they are explicit as to what they intend the overall consequence is a degree of confusion. For example, in his paper "Moral nativism: A sceptical response," Kim Sterelny is careful to explain that he is skeptical of the *developmental* nativist thesis: He allows that "there is a plausible ... case for the idea that moral cognition in an adaptation," but adds that "even if that is right, it does not follow that this capacity is innate" (2010: 280). If such qualifying comments are overlooked, however, then one might gain the impression that Sterelny is opposed to those who advocate moral nativism as an *adaptational* hypothesis, when in fact they may be in complete agreement.

In the present context, focused on the relation between evolution and ethics, it is some version of the adaptational hypothesis that is under consideration. That is not to say that the developmental version of moral nativism is irrelevant—its truth or falsity may have important bearing on the evolutionary thesis—but it is the adaptational hypothesis that is our target. But this doesn't quite settle what is meant by "innate" here, for one may wonder what kind of *adaptation* is intended. Adaptations are traits that emerge and persist in a population through a process of selection: because they are transmitted from parents to offspring and provide their bearers with a reproductive advantage. But this central Darwinian idea is neutral concerning the mechanisms of inheritance. The widespread assumption that genetic transmission is the only mechanism that counts is mistaken; traits passed on through cultural channels can just as well count (by orthodox Darwinian standards) as adaptations. (See Boyd & Richerson 1985; Richerson & Boyd 2005.) Hence, one could legitimately claim that human morality is a Darwinian adaptation (thus advocating moral nativism) while also maintaining morality to be an entirely cultural phenomenon. Nevertheless, this is not the way the matter is generally seen at present. It seems fair to say that nearly everyone who discusses moral nativism as an evolutionary hypothesis has in mind the view that morality is a *biological* adaptation transmitted via genetic mechanisms, and this is what I shall assume in what follows.

Let us now turn to the nature of the trait, which I have been calling simply "human morality." The earlier introductory outline specified that this phrase denotes the capacity to make moral judgments. But further indeterminacy arises, since it is not at all clear what makes a judgment a *moral* judgment, and even if that were settled there remain a number of distinct hypotheses concerning what aspect of moral judgment might be an adaptation.

What is a moral judgment?

Let us start by distinguishing moral nativism from what might be called "altruism nativism." Here I use "altruism" in the vernacular sense, to denote certain motivations and actions performed from those motivations. An altruistic act is one that is done with the ultimate goal of benefiting another. Whether one *succeeds* in benefiting the other is irrelevant to whether the action is altruistic. An act that is done in order to benefit another but only because in

providing that benefit one in turn profits *oneself*, is not an altruistic act. Since altruism in this sense requires the cognitive capacity to conceive of *oneself* and of *others*, it (and its contrasting sense of *selfishness*) can be sensibly ascribed only to cognitively sophisticated creatures. Indeed, humans are the only undisputable case.

It is important to distinguish this psychological sense of altruism from evolutionary altruism. The latter is a characteristic of traits that have been selected for because they decrease the reproductive fitness of their bearer while increasing the fitness of others. Evolutionary altruism has nothing to do with psychological states; it is something that can sensibly be ascribed to plants or protozoa. Many cooperative behaviors in nature (social grooming, food sharing, group hunting, etc.) are casually referred to as instances of "evolutionary altruism" when in fact this is questionable. It depends on whether one understands "fitness" to mean (A) fitness over the life of the organism, and (B) reproduction that is achieved indirectly as well as directly. A primate that grooms another because the favor will be repaid in the future is not sacrificing its fitness if we measure fitness over its life time. A wild turkey that foregoes seeking mates in order to help its brother in his mating displays is not sacrificing its fitness if we measure fitness in a way that includes its indirect (inclusive) contributions to the replication of its genes. With these restrictions made, one may wonder whether genuine evolutionary altruism is even a possibility. Elliot Sober and David Sloan Wilson (1998) argue that it is a possibility (and, indeed, a reality) so long as populations are structured in such a way that the altruistic trait, though detracting from the reproductive fitness of its bearer relative to other group members, nevertheless provides reproductive advantage to that group relative to other groups.

Cooperation is often the means to reproductive success. When natural selection favors cooperation—whether it is evolutionarily altruistic or evolutionarily selfish—then proximate mechanisms governing the behavior will be necessary. These mechanisms will take a myriad of forms in nature, for they include the workings of everything from algae to zebras. For creatures with the cognitive wherewithal, there is no reason why natural selection may not plump for *altruism*, in the aforementioned psychological sense, as a capacity to encourage forms of cooperation. For example, a parent who simply cares directly for the welfare of his or her children is, arguably, moved by a more reliable and frugal mechanism than the psychologically selfish parent who must perceive a contribution that the children's welfare makes to his or her own wellbeing.[1] (See Sober 2000; see Stich 2007 for criticism.) Though the argument seems strongest when applied to parent-offspring bonds, there is nothing to prevent its being employed in reference to other kinds of cooperative relations that were adaptive in the ancestral environment, such as reciprocal and mutually beneficial interactions among non-kin.[2] Such considerations form a plausible basis for supposing that nativism concerning human psychological altruism might be true.

---

[1] It is vital to underline that although the parent's caring behavior may be evolutionarily selfish, this in no way implies that it in some sense "really" psychologically selfish.

[2] In a reciprocal exchange, the proportionally greater payback one receives for one's efforts may explain why the capacities to engage in such behavior were selected for, but it is an error to assume that this reveals anything about one's psychological motivations. Of course, in humans reciprocal relations are often governed by pragmatic thoughts of payback, but they need not be; there is no a priori reason for excluding the possibility of evolution preferring to encourage such behavior by granting parties (possibly conditional) altruistic concerns for their exchange partners.

But considerations in favor of altruism nativism fall short of considerations in favor of moral nativism. To see why, one need only reflect on the fact that a person can be moved by altruistic sentiment while making no moral judgment whatsoever. A parent's powerful motivation to care for his or her children is typically governed simply by the raw *desire* to do so, not by any sense of duty. This is not to say that parents would deny the presence of parental duties, but rather that they have entrenched feelings of affection for their children independently of any moral judgments. Acting in a helpful manner because one wants to is simply not the same as acting in a helpful manner because one judges that one ought to. It seems that we can, in fact, imagine a social species that evolves with strong feelings of altruistic concern for each other but whose idiosyncratic neurology renders them entirely lacking in the conceptual prerequisites necessary for making moral judgments. Such creatures cooperate because they love each other, but they have no capacity to think of cooperation as a morally desirable or morally required activity. Altruism nativism would be true of such creatures, but moral nativism would not be.

Perhaps such imaginary creatures are not a million miles away from some of our primate cousins. Primatologist Frans de Waal has argued that apes and some monkeys have many of the "building blocks" of morality, such as empathy, a sense of fairness, and the capacity to engage in consolation behavior (1992, 1996, 2006). But while these traits might suffice for what de Waal calls "a sense of social regularity" (1992: 242), he admits that they fall short of anything truly deserving the name "a moral sense." But what other building blocks need to be added? What is it that the chimpanzee lacks?

A tempting (though contentious) answer is that the chimpanzee lacks certain *cognitive* resources. A chimp that uses its canine teeth during an in-group dispute incurs an angry response from the other group members, even those not directly involved in the conflict (de Waal 1992: 257). Yet it is unlikely that they think of the action as *justifying* the reprimand; it is doubtful that they can conceive of the action as wrong in a cool-headed way, in the absence of angry arousal. The intensity of the punitive response will depend entirely on the contingent intensity of the immediate arousal; it will not derive from deliberation concerning what level of reprimand is fair. The transgressing chimp may fear the negative response from the group, but it seems improbable that he can grasp that he might *deserve* that response—a conceptual achievement necessary if *guilt* is to be ascribed.

These thoughts are little more than gestures in a certain direction, but they do at least accomplish that: to point toward a promising research program. According to this line of thinking, moral judgment requires the capacity to deploy concepts like *desert*, *wrongness*, and *justification*. An additional common idea is that moral judgments are distinguished from other kinds of normative judgment (such as concerning prudential and conventional norms) by their being imbued with a certain distinctive practical authority. Moral prescriptions are those that one *has to* follow whether one likes it or not; moral imperatives, in other words, are not pieces of advice on how to satisfy one's ends. (See Kant 1783/1985; Joyce 2001.) If the practical authority of moral imperatives doesn't derive from their conducing to one's ends, then whence does it derive? Arguably not from any human authority. Moral norms are often contrasted with conventional norms on the grounds that the latter are allowed to depend on the decree of some authoritative body (be it monarchs, teachers, God, or collective opinion), whereas the former are taken to be transcendent of any institution. (See Smetana 1993; Nucci

2001; Turiel 2002.) (Note that we are discussing the authority with which moral norms are imbued; whether moral norms actually *have* this authority is another matter entirely.)

The preceding is admittedly all rather vague and is certainly controversial. This is one of the difficulties with the debate over moral nativism: People argue at length over the origins of a trait but often have only a slippery grasp of what the trait is, or, even when they have a crisp idea in mind, their stipulative construal of the trait differs from that of others. Until there is agreement as to which trait is under discussion, confident claims about its historical source are rather premature—an admonition applying as much to the anti-nativist as the nativist. Nevertheless, with this important hesitation noted, there is nothing to stop one signing a promissory note to attempt to settle the matter in the future, and proceeding further with the discussion in a provisional fashion.

Different moral nativisms:

Even if we knew precisely what a moral judgment is (or at least made a collective decision in this regard), there would still be some question as to what moral nativism is. At one extreme is the nativist claim that no particular moral judgment is innate but that the general faculty for making these judgments is innate. One way of conceiving of this view is that the human brain comes prewired with moral *concepts* (concepts like *wrongness*, *goodness*, etc.) and the social environment teaches the individual to which items to attach these concepts. Thus one environment may teach the child to judge that killing foreigners is morally wrong, while another environment may teach another child to judge such behavior as acceptable or even admirable. This variation would be no mark against the kind of nativism under discussion, for both parties would equally exhibit the trait in question: the tendency to make moral judgments.

Another kind of moral nativism holds that the concepts provided by the moral faculty come with biases toward certain subject matters. Psychologists Jon Haidt and Craig Joseph (2004), for example, argue that a comprehensive survey of cross-cultural data reveals that the moral sense comes prepared to manage certain broad domains: actions producing harm, relations concerning fairness and exchanges, values pertaining to social hierarchy, and regulations surrounding certain bodily matters (such as menstruation, sex, corpses, etc.). According to this view, humans will find certain moral systems easier to learn than other systems. (See also Haidt & Bjorkland 2008.)

Another kind of nativism will hold that specific moral judgments are biological adaptations. Clearly there is nothing to be said for the view that *all* moral judgments are biological adaptations (for consider a moral judgment about that particular lie that Fred told last Tuesday). But perhaps for certain general complete judgments, concerning matters that were present in the prehistoric environment, the nativist claim might be upheld. Jesse Prinz (2009), for example, discusses nativism for the judgments "Don't harm innocent people," "Respect and obey authorities," and "Incest is prohibited." (Prinz argues emphatically against this nativism, but the fact that he finds it necessary to do so shows that it must be a viable hypothesis.)

Perhaps a few broad abstract moral principles are prewired into the human brain, and in addition there are certain innate parameters, such that the social environment sets with which

parameter the principles combine in order to produce a full range of specific moral judgments. This is the position of John Mikhail (2011) and of Marc Hauser (2006), drawing inspiration from Noam Chomsky's view on linguistic nativism.

This significant variation in different forms of moral nativism is important to appreciate, since some versions may be more plausible than others, evidence mustered in support of one version may not serve to support others, and sound arguments against one version may not trouble others.

Why might moral thinking have been adaptive?

Any evolutionary moral nativist owes us an account of why moral judgment (however it is to be defined) was reproductively useful—what the "adaptive problem" was to which the trait was the solution. Such an account need not be pure speculation; it can to some extent be based on careful empirical observation of the ways in which moral thinking continues to impact upon our practical lives. The provision of such an account does not show that nativism is true, of course, but nevertheless contributes to its plausibility.

On this matter (as with every other element of moral nativism we have thus far discussed) there are various options for the nativist. One important distinction concerns *whose* reproductive fitness is enhanced. It is natural to assume that one should examine the profits for the individual who bears the moral trait. If, however, one is inclined to suspect that the moral faculty evolved via a process of group selection—of the kind associated above with Sober and Wilson's views on evolutionary altruism—then it is the reproductively relevant profits to the bearer's *group* that are pertinent. Another distinction concerns whether one is considering *self-oriented* or *other-oriented* moral judgments. It is possible that one kind of moral judgment was primarily adaptive and the other emerged subsequently as a kind of useful by-product. Darwin, for example, in *The Descent of Man* frequently identifies the moral sense with the *conscience*, indicating a focus on self-directed moral judgments; whereas Edvard Westermarck offers the reverse view: that self-oriented moral assessment is attained only "through a prior critique upon our fellow-men" (1906: 123).

Whatever options are taken in these respects, most moral nativists suggest that moral judgment was adaptive because it in some manner encouraged ancestral cooperation. It is plausible that in certain contexts thinking of a cooperative venture as *obligatory*, say—such that defecting on the deal will leave one feeling that punishment is not merely risked but *deserved*—will strengthen one's resolve to perform the action. (See Joyce 2006, chapter 4.) Needless to say, this trait could be disastrous if attached to maladaptive behavior (e.g., over-cooperating), so the "moralization" will need to be sensitive to environmental variables. According to this view, the evolutionary function of morality is to act as a kind of motivation enhancer, to offset the possibilities of practical self-sabotage (e.g., succumbing to temptations of immediate gratification). These possibilities may arise because of the operations of other psychological faculties (e.g., our capacity to calculate for self-gain) which are generally adaptive but occasionally steer us toward suboptimal decisions.

It would be a mistake, however, to see the benefit of the moralization of certain cooperative behaviors simply in terms of a private little mental nudge. Such a view fails to appreciate the social dynamics of moral assessment. Moral evaluation is something that can

be employed publicly to condemn an action or a person, to justify a punishment, and to demand others to participate in the punitive response. Moral values can be seen to be *shared*, thus enhancing social cohesion even when the moral value itself does not pertain to any cooperative behavior (but concerns, say, the treatment of food). Thus moral judgments can function usefully not just as personal commitments, but can be signaled in a way that makes them potential interpersonal commitments. The fact that abiding by moral standards generally involves foregoing short-term profits means that morality can function well as a *costly* signaling device. When choosing partners for a mutually beneficial cooperative venture, it makes sense to prefer those who can *honestly* signal their willingness to participate. And making signals *costly* is a way of making them honest, for a sufficiently expensive signal costs the signaler more than the profits that might be reaped through dishonesty. (See Zahavi 1977; Noë 2001.) Thus, if one's flourishing or very survival depends on being chosen in cooperative ventures (whether it be as a mate or as a member of a hunting party), it may be adaptive to signal in a costly way one's social virtues.

In accounting for the adaptiveness of moral judgments, the moral nativist may therefore choose to downplay their role as motivation enhancers and instead highlight their potential as grounding public signals. Psychologist Geoffrey Miller (2000, 2007), for example, explores how morality serves as a signaling device in human mate selection, and thus argues for an unusual kind of moral nativism according to which the moral sense evolved through a process of sexual selection. Economist Robert Frank (1988) argues that moral emotions evolved as signaling devices—that the facial expressions accompanying emotions, which in large part lie outside our autonomous control, are nature's way of guaranteeing a degree of honesty in our social interactions, in a manner that gave our ancestors who bore this trait a social advantage over the competition, which translated into a reproductive advantage.

It is not necessary to see the adaptiveness of moral judgment in terms of its social role. Daniel Dennett (1995) suggests that moral thinking serves as a "conversation stopper": a no-questions-asked practical consideration that silences any further deliberation on the matter in question. In certain circumstances, we humans are too clever for our own good; we have the capacity to keep worrying about costs and benefits, possible consequences, etc., to such an extent that the very labor of calculating the optimal course of action ensures a suboptimal outcome. Hence, absolute categorical rules and values can play a very useful role in deliberations, bringing them to an endpoint on an outcome that is probably satisfactory. There is nothing to say that these rules and values must concern social matters (like *Don't kill your neighbors*); they may fulfill this function just as well if they concern entirely private affairs (like *Don't risk your own life for trivial gains*).

While the moral nativist certainly needs to say something about how moral thinking was adaptive for our ancestors, in one sense this is the least challenging aspect of making the case, since nearly everyone—anti-nativists included—accepts that moral thinking serves some useful social purposes. Of course, moral thinking can lead to disasters as well, so there is a substantive question as to whether it is all-things-considered a good thing. (For doubters, see Hinkfuss 1987; Garner 2010; and, in certain moods, Nietzsche 1887/1994.) But the moral nativist can expect to gain agreement from at least most of the opposition that morality was/is generally useful (if not to the bearer of the trait, then to his/her group), and this usefulness is likely to translate in some manner into a plausible case for reproductive advantage. The

harder challenge for the moral nativist is to find evidence for the plausible hypothesis.

Evidence for moral nativism:

Perhaps the only claim that can be made confidently about the status of evidence either for or against moral nativism is that the whole debate is in a state of disarray. There is fundamental lack of consensus concerning what would even count, in general, as evidence that a psychological trait is an adaptation; and therefore it comes as no surprise that there is much confusion when it comes to the trait of moral judgment, about which, as we have seen, there is no fixed view as to what the trait even is. (Discussions of empirical methodology regarding psychological adaptations include Ketelaar & Ellis 2000; Conway & Schaller 2002; Confer et al. 2010.)

Once a version of the moral nativist hypothesis has been focused upon, testing must proceed as with any empirical hypothesis: Predictions of the hypothesis must be identified, the evidence of whether these predictions obtain must be gathered, and alternative explanations of the data must be examined. While such truisms can be agreed upon, when it comes to testing an adaptationist hypothesis there is trouble at every step.

It is often assumed, for example, that one of the predictions of nativism is the universal (or near-universal) presence of the trait in human populations. Thus, much energy is expended on cross-cultural studies, both by nativists trying to demonstrate universality and by anti-nativists trying to provide counter-examples. If the hypothesis under discussion is nativism *in the developmental sense* mentioned earlier, then evidence pertaining to universality can be effective, for moral nativism in that sense is the claim that moral judgment is a trait whose emergence is reliable in the face of environmental perturbation; thus to observe morality emerge ubiquitously in a highly varied range of developmental settings would constitute solid confirming evidence. But the relation between *evolutionary* nativism and universality is much more complicated, to say the least.

It was noted earlier that adaptations may require specific environmental input. This may be in the form of external "cues" that trigger development. If these cues were reliably present during the period of selection, then there would be no pressure to make the emergence of the adaptive trait assured in the absence of the input (see Griffiths 2002: 74-5). Yet the modern environment differs in countless ways from the ancestral one, raising the possibility that certain cues are now given unreliably, in distorted form, too often, or not at all. Alternatively, the developmental process may be "expecting" differential cues, signaling the need to take one developmental trajectory or another. For example, whether male European earwigs (*F. auricularia*) develop long or short forceps, which determine distinct mating strategies, (the two forms being so different that they were once considered distinct species), depends on local population density (Tomkins & Brown 2004). In the latter case, what is selected for is a mechanism that executes a conditional strategy depending on environmental variables. In the former case, even the conditional mechanism might not develop due to the absence of crucial environmental cues during development. (See Buller 2006: chapter 2.) In either case, we should not expect to see a universal phenotype associated with the adaptation.

The reverse implication also fails. Even if we do find a universally present trait, it is very difficult to exclude alternative non-adaptationist hypotheses. The trait might not be itself an

adaptation but rather a "spandrel" that piggybacks on other adaptations. Male nipples are universally found (in males) but are not an adaptation but a spandrel (Gould 1992). In the case of moral judgment, a conspicuous alternative hypothesis is that moral judgment serves no evolutionary function but is an inevitable accompaniment of other psychological faculties that are adaptations. If these other adaptations enjoy universal manifestation, then so too will the moral capacity. Alternatively, even if moral judgment does not *inevitably* accompany adaptational traits, these traits might make moral judgment possible for the human mind, and moral systems might simply be a fairly obvious invention for groups of humans living together. This is what Dennett calls a "Good Trick": a solution to a recurrent problem that is sufficiently straightforward for creatures with our array of sophisticated cognitive and emotional capacities that one can expect it to be struck upon pretty much everywhere (Dennett 1995: 77-8, 485-87). The observation of universality alone does not favor the nativist hypothesis over either the spandrel or Good Trick hypothesis.

The poverty of the stimulus (POS) argument is another kind of evidence in favor of nativism that is more problematic than is often realized. The structure of the argument comes, of course, from the debate over nativist explanations of human linguistic abilities (see Chomsky 1967, 1987/1990), where the POS argument is widely judged to be triumphant in establishing some form of nativism. According to this argument, the capacities evident in language use emerge in a manner that far outstrips the information that is available in the learning environment. What are the prospects of a *moral* POS? The answer is that there are challenges at every step of the argument.

First of all, it is not clear to what extent the hypothesis that moral judgment is a biological adaptation predicts that the trait will emerge in conditions of impoverished stimuli. Some versions of moral nativism described earlier require a substantial amount of *learning* in the acquisition process. According to one version, for example, what biological selection provides is a mechanism that makes possible a particular special kind of learning (i.e., the acquisition of moral norms). If such a version is under scrutiny, it would be a distraction to examine how a child acquires a full moral judgment, like *Stealing toys is wrong*, for we can all agree that the child is taught this by his/her parents. Rather, what one should be asking is how the child comes by the concept *wrongness* in the first place. But even here, as we have seen, the hypothesis that the possession of this concept is a biological adaptation is consistent with its developmental emergence requiring specific environmental input.

Second, it is debatable to what extent the observable data reveal that the moral acquisition process unfolds in conditions of poor stimuli. A lot of research is understandably focused on how *early* traits appear in development. Psychologist Paul Bloom and colleagues have found evidence that infants as young as three months preferentially discriminate others on the basis of their social behavior toward third parties (Hamlin, Wynn & Bloom 2007, 2010.) Nobody maintains that these infants are making full-blown moral judgments, of course; but the full-blown trait in which we are interested appears surprisingly early. Following the lead of Elliot Turiel, several developmental psychologists have investigated the emergence of the capacity to distinguish moral from conventional norms, finding evidence of the capacity in children yet to turn three (Smetana & Braeges 1990). Yet even here it is questionable to what extent this early emergence occurs in conditions of impoverished stimuli. By the age of three the child has been exposed to an enormous amount

of unrelenting instruction from its parents, and has usually been able to observe a broad variety of social interactions. Shaun Nichols reminds us that "the child is exposed to lots of admonitions and instruction in the normative domain. Parents and teachers are constantly telling kids what shouldn't be done" (2005: 358). Sterelny makes a similar observation:

> The narrative life of a community—the stock of stories, songs, myths and tales to which children are exposed—is full of information about the actions to be admired and to be deplored. Young children's stories include many moral fables: stories of virtue, of right action and motivation rewarded; of vice punished. So their narrative world is richly populated with moral examples. (2010: 289)

When it comes to assessing how rich or poor is the child's learning environment, what is crucial (again) is identifying just which trait it is whose emergence is under discussion. But even making this decision leaves many challenges. Suppose, for example, that the target hypothesis concerns how a child acquires the basic moral concepts (*wrongness*, etc.) in the first place. The problem is that we do not know even in principle what combination of endogenous and exogenous factors are necessary or sufficient for such a concept to become available in a developing human brain. The puzzle can be put informally as follows: Let us grant the developing child all the careful instruction, all the scaffolded learning, all the varied experiences, all the trial-and-error social interactions, all the exposure to moral tales and exemplars, all the coordinated rewards and punishments, and so on, that one cares to—let us, in short, allow the stimulus to be as rich and varied as one likes; the challenge remains: What psychological mechanisms must be present in order for all this environmental input to result in the emergence of the capacity to employ a moral concept? Since there appears to be no consensus on how this question should be answered, the possibility remains open that one or more of the necessary mechanisms must be *dedicated* to this particular kind of acquisition, in the sense that they have been forged by biological selection for the task.

A third difficulty for a moral POS argument is that even if the adaptational hypothesis did predict that moral development would occur with impoverished stimuli, and even if we had solid evidence that moral development *does* have this characteristic, the moral nativist still faces the challenge of excluding alternative hypotheses, and this is no easy matter. It has often been noted that the conclusion of the POS argument is negative—it establishes one means (i.e., learning) by which the psychological capacities in question have not come about (see Lawrence & Margolis 2001; Nichols 2005). The success of such an argument might support nativism of a developmental sort, but it does not suffice for nativism in the evolutionary adaptational sense in which we are interested here. A trait may become manifest prior to any learning, indicating that it is in some sense "non-acquired," but this would not suffice to show that it is an adaptation. The prominent alternative hypothesis that is not excluded is that the trait is a spandrel. A spandrel that piggybacks on other traits, which are adaptations, will emerge just as they do. If the adaptational traits are expressed early, in a predictable sequence, universally, and in environments of impoverished stimuli, then so too will the spandrel trait.

Alternative by-product hypotheses:

Several opponents to moral anti-nativism have offered what amount to alternative spandrel hypotheses—or at least *by-product* hypotheses.[3] The standard approach is to offer some psychological faculties that are not specifically *moral*, then attempt to show that the capacity for moral judgment could (or would) emerge from these traits. Darwin himself can be interpreted in this way. He writes that "any animal whatever, endowed with well-marked social instincts, … would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well, or nearly as well developed, as in man" (1879/2004, 120-1). These "intellectual powers" include a good memory, language, and the capacity to form habits of thinking and acting. Thus it seems that the moral faculty, for Darwin, is a spandrel of other psychological capacities. Even regarding the social instincts themselves he is undecided as to whether they are adaptations or by-products, writing that it is "impossible to decide in many cases whether certain social instincts have been acquired through natural selection, or are the indirect result of other instincts and faculties" (1879/2004, 130).

Modern anti-nativists do not concur regarding Darwin's exact ingredients (the one who comes closest is Francisco Ayala (2010)), but the pattern of argument remains the same. Prinz (2008), for example, tries to build moral judgment out of non-moral preferences (e.g., for fair reciprocal exchanges, against incest), meta-emotions (emotions that take other emotions as their subject), perspective taking (allowing for third party concern), and non-moral emotions such as sadness and anger. Nichols (2005) allows that nativism might be true of the capacity to use non-hypothetical imperatives (i.e., imperatives that do not depend for their legitimacy on the addressee having certain ends), but he rightly notes that moral imperatives are but a subset of non-hypothetical imperatives (e.g., "Do not move your rook diagonally" is a non-hypothetical but non-moral imperative). Nichols adds that this capacity might combine with another innate tendency—namely, an affective mechanism that responds to suffering in others—to explain why the domain of moral normativity gets singled out as salient, resonant, and memorable. He writes:

> [B]oth of the mechanisms that I've suggested contribute to moral judgment might well be adaptations. However, it is distinctly less plausible that the capacity for core moral judgment itself is an adaptation. It's more likely that core moral judgment emerges as a kind of byproduct of (*inter alia*) the innate affective and innate rule comprehension mechanisms. (2005, 369)

Such offerings must be assessed on a case by case basis. The key question to examine is whether the non-moral ingredients offered really do suffice to account for the capacity to make moral judgments. Simply asserting that they do is an easy undertaking, but a proper critical examination of the claim may be a complex matter requiring various kinds of empirical investigation. One factor which might confound the debate is the possibility that the very notion of *moral judgment* is to a greater or lesser extent indeterminate, such that the anti-nativist's ingredients suffice to explain moral judgment in some attenuated form while not accounting for a more full-blooded conception of the phenomenon (see Joyce 2012).

---

[3] Following Ben Fraser (2010), I take spandrels to be traits that *necessarily* accompany adaptational mechanisms (like Stephen Jay Gould's original architectural spandrels, which are *unavoidable* features of a cathedral built with such-and-such a functional design); whereas *by-products* are traits that the adaptational mechanisms make possible but not inevitable. For the original metaphor, see Gould & Lewontin 1979.

Any proponent of a by-product hypothesis needs to be careful that it really is a by-product hypothesis that is being articulated rather than a useful clarification of a nativist hypothesis. Let me explain using Nichols' suggestion as an illustration. On one reading (the one he intends), Nichols, if successful, will have accounted for the trait of moral judgment by appeal to other evolved mechanisms. But on an alternative interpretation, Nichols has simply succeeded in providing some extra detail for the nativist case. Notice that how we enumerate psychological "mechanisms" is always to some extent dependent on our focus; what from one theoretical perspective is a single mechanism is, from another perspective, a group of mechanisms, each of which is made up of a group of mechanisms. Thus, if humans do have an innate "moral faculty," one should not expect a monolithic entity that performs moral judgment (arguably not even a coherent idea), but rather a group of innate sub-faculties. Perhaps, then, Nichols has simply succeeded in delineating two of those sub-faculties.

The issue turns on whether these two mechanisms are "supposed" (from an evolutionary perspective) to work in tandem. On the one hand, we can picture two biologically evolved mechanisms ticking along, interacting in certain ways, and the consequent joint output having characteristics that are wholly accidental. In this case, the output should be considered a by-product or spandrel. On the other hand, we can picture this output having some reproductive relevance, and thus natural selection taking an interest in their interaction, tinkering with the human genome in order to encourage the interaction. In this case, at a certain point we must cease to speak of the output as a "byproduct" of two innate mechanisms, and rather speak of it as issuing from a dedicated mechanism (which consists of two preadaptional sub-mechanisms). According to this latter account, one might call the moral faculty an *exaptation*—a result entirely acceptable (and, indeed, unsurprising) to the moral nativist.[4]

Even if the moral anti-nativist succeeds in articulating a clear by-product hypothesis, and avoids the charge that what has really been accomplished is a delineation of the sub-mechanisms comprising the adaptational moral faculty, this would not show that moral nativism is false. I know of no good reason for endorsing *methodological anti-nativism*, according to which if a non-nativist explanation of a trait is possible, then it is to be preferred. Such a principle is at best strictly optional, and at worst entirely question-begging. (Of course, methodological nativism would be just as bad; but why presuppose "methodological *anything*-ism" in this respect?) Even if one can conceive of a way that general learning mechanisms and non-moral faculties could suffice for moral judgment, the latter competency might nevertheless as a matter of fact be brought about by dedicated innate mechanisms.

---

[4] I introduce the term "exaptation" with some hesitancy, since a great deal of confusion surrounds it. (See Buss et al. 1998 for excellent analysis.) Biological natural selection, as we have seen, can produce both adaptations and spandrels/by-products. Both kinds of trait can be coopted for new uses, and when they are then they become *exaptations*, to use the term introduced by Gould (Gould & Vrba 1982; Gould 1991). Gould is unclear about what processes can perform this "coopting," but it seems reasonable to maintain that when the process is natural selection then the exaptation is also an adaptation. Feathers, for example, possibly evolved originally as devices for thermo-regulation; this was their adaptive function. Gradually, however, in certain organisms they were coopted for a new function: they aided gliding and ultimately flight. While this permits one to describe modern avian feathers as an "exaptation," it seems unnecessarily counterintuitive to assume that it thereby excludes referring to them as an "adaptation." Dennett has complained that really *all* adaptations are exaptations (1995: 281), but David Buss and colleagues have plausibly argued that there is some explanatory benefit in maintaining a distinction (see Buss et al. 1998: 542). However, there is no need to see exaptation and adaptation as contrary categories; the distinction can be maintained perfectly well if one classes the former as a proper subset of the latter. See Fraser 2010 who usefully discusses the place of exaptation in the moral nativist debate.

Natural selection might, for example, have reasons for favoring a more reliable and faster acquisition process over empirical learning, even if acquisition via learning remains possible. Empirical scrutiny of the actual acquisition process might reveal that the anti-nativist's plausible hypothesis is simply not how the process occurs.

In the foregoing I have not attempted to put forward evidence for or against moral nativism, but rather draw attention to the complexities of the matter. The fundamental message is that nativist hypotheses are orthodox empirical claims and must be treated as such. As with any area of science, there is a place for imaginative speculation and weighing matters in terms of what seems most plausible. Of course, if anyone thought that a plausible bit of speculation could be considered *true*, then Gould's famous allegation that this amounts to no more than a *Just So Story* would be fair. However, it is far from clear that any serious moral nativist does make so crass an error; most appear acutely aware that their hypotheses require evidential support and that the empirical project is still in its infancy.

3. Implications for moral philosophy

Ever since the ancient Greeks it has been possible to discern a broad dichotomy in moral philosophy: between those views that see the moral realm as somehow transcendental and objective and those views that see morality as dependent on embodied humans that are products of nature. Those who favor the latter perspective will find it hard to resist the thought that discoveries concerning the Darwinian processes that shaped our biological nature must have some relevance to human morality. But what kind of relevance? Possible answers are too numerous to be cataloged here.

The first philosopher to attempt to use Darwinian theory to justify particular moral decisions was Herbert Spencer. Spencer championed a progressive view of evolution: that the inevitable outcome of natural selection is superiority of organisms (more complex, smarter, more cooperative, more harmonious): "The conduct to which we apply the name good, is the relatively *more evolved* conduct" (1879/2011: 26). From this dubious basis, he argued that certain laissez-faire social policies followed, that it is often the role of the state to let matters run their natural course with respect to society's "unhealthy, imbecile, slow, vacillating, faithless members" (Spencer 1851: 324).[5] Arguably, Spencer's error here derives from confusion over what it meant by Darwinian *fitness*—hardly a surprise when it is recalled that Spencer coined the phrase "survival of the fittest": a problematic expression that has often been accused of tautological vacuity. Organisms that are *fitter* than their competitors in the Darwinian sense are simply those that replicate their genes more successfully. But these fitter organisms need not be fitter in the intuitive vernacular sense. They may be less complex, less intelligent, and less cooperative; they may be more unhealthy and more prone to imbecility and faithlessness (if, for example, these are the traits strongly preferred by potential mates). Spencer slides between these different senses of "fitness," and thus fails to see that when the welfare state chooses to aid a starving pauper, say, it is not a case of an evolutionarily unfit

---

[5] In reality, this is but one thread of Spencer's changing and complex views on moral philosophy. See Weinstein 1998.

organism being unnaturally propped up, but rather the organism in question being (in all probability) rendered evolutionarily fitter.

Contemporary philosophers have pursued various strategies for using evolutionary thinking to vindicate morality, either regarding particular practical decisions or more generally. (See Richards 1986; Rottschaefer & Martinsen 1990; Dennett 1995.) Here I will focus briefly on just one family of arguments that try to find moral normativity in the evolutionary process.

If moral judgment is an adaptation, then we can legitimately speak of its "usefulness," from which one might draw a conclusion concerning *justification* (Campbell 1996). Moral nativism also allows one to speak of morality's having a biological function, from which one might draw a conclusion concerning what morality is "supposed" to do, and of the moral sense operating "well" or "poorly" (Kitcher 2011). Such positive-sounding conclusions for morality are, however, not of the kind in which a moral philosopher is typically interested. From the fact that a trait is the product of biological natural selection we can draw two conclusions concerning its usefulness: first, that it *was* useful during the relevant period of selection; second, that it was useful for the replication of genetic material. It doesn't follow that it *is* useful now, nor that it is useful for any ends that an individual might have that do not concern genetic proliferation, of which surely we have an abundance. Perhaps even more important to note is that any such justification is entirely *instrumental*—not the kind of *epistemological* justification with which metaethicists are usually concerned.

From the fact that a trait's having a biological function gives license to an array of normative-seeming language (concerning what it is "supposed" to do, what is involved in its functioning "well," and so forth), no real practical authority arises. By analogy with artifactual functions: If an archeologist unearths an ancient taro pounder she is under not a glimmer of obligation to use the item for pounding taro. Similarly, consider a character like Hume's sensible knave, who observes the norms of morality only on the grounds of expediency but who doesn't hesitate to take full advantage of any exception that arises, even if it is at great cost to others. It may well be that in choosing to break an inconvenient promise the knave is using his moral sense in a way that it is not "supposed" to be used, or perhaps failing to use it at all, but this fact alone—the fact that a biological function is going unfulfilled—doesn't seem sufficient to underwrite our assessment that he is acting in a morally prohibited manner.

This represents a serious problem for virtue ethicists who look to the Ancient Greeks for inspiration but hope to appeal to Darwinian teleology to replace the bankrupt Aristotelian teleological worldview (see Casebeer 2003; Brown 2008). Evolutionary functions alone simply do not provide the action-guiding normativity with which virtue ethicists hope to imbue the virtues.

Recent years have seen explored a very different agenda concerning deriving metaethical conclusions from evolutionary data: the contention that moral nativism *undermines* morality. "Undermine" here is an intentionally vague term, covering different possibilities:

1. Moral judgments are unjustified.
2. Moral judgments are unjustified and irredeemably so.
3. Moral objectivity is an illusion; therefore all objective moral judgments are false.

4.   All moral judgments are false.

Michael Ruse often couches his evolutionary moral skepticism in terms of lack of justification ("What kind of metaethical justification can one give for [moral] claims?" (Ruse 2006: 20)), but it would seem that his ultimate ambition is to establish a claim of falsehood ("Morality is a collective illusion foisted upon us by our genes" (Ruse 1986: 253)). Sometimes Ruse seems to be aiming to establish only thesis 3, but when one couples 3 with his explicit claims that objectivity is part of the *meaning* of morality—e.g., "Ethics is subjective, but its meaning is objective" (Ruse 2006: 22; see also Ruse 2009: 507)—one can draw the conclusion that morality *tout court* is false. In other words, one might explicitly claim only that *moral objectivity* is a flawed notion, but if one also maintains that morality is necessarily objective (conceptually speaking), then one is in effect asserting that morality *per se* is flawed. (This, at least, is how I shall interpret Ruse here, if only to have a concrete advocate of thesis 4 before us.)

What role does moral nativism play in Ruse's debunking argument? Not only does he endorse moral nativism, but he argues that it explains why we imbue our moral claims with *objectivity*, since this is a crucial element in his account of why moral thinking was adaptive to our ancestors. Ruse then deploys a parsimony principle to establish skepticism. All that needs to be explained, he thinks, has been explained. There is simply no need to go further and posit the realm of moral properties that is necessary to make moral judgments *true*. His favored analogy is with explaining the popularity of spiritualism after World War 1. By appealing to psychological and sociological factors we can explain all we could want to about why a grieving mother might come to believe that her dead son spoke to her at a séance, without having to posit the realm of souls-talking-from-beyond-the-grave that would be necessary to make her belief *true*.

The analogy is not entirely persuasive, however. Whereas it is clear that what is needed to render the spiritualistic belief true is a whole layer of extra spooky ontology in the universe, this is not so obvious in the moral case. Most modern moral realists attempt to account for moral properties on the basis of antecedently accepted ontological categories. (Hedonic utilitarianism, for example, proposes to build moral obligation out of happiness plus the causal relation of being productive of happiness.) Thus it is not obvious that an application of Ockham's Razor is apposite in the moral case.

A principle of parsimony also plays an important role in Sharon Street's attempt at an evolutionary undermining of morality (Street 2006). Her ambition is weaker than Ruse's, since she aims to use moral nativism to establish only thesis 3. Her opponent is moral realism (understood as including an endorsement of objectivity), and since she allows that some kind of moral constructivism might survive the argument it is evident that she does not endorse the claim that objectivity is an essential part of the meaning of moral terms.

She proceeds by observing that moral nativism presents the moral realist with a dilemma: Either (i) there is no causal connection between the (supposed) realm of objective moral facts and the moral beliefs with which natural selection has endowed us—in which case there is very little chance that our moral beliefs match the facts—or (ii) there *is* a causal connection, namely that our moral sense has been designed to track the objective moral facts. The problem with the second horn of the dilemma, according to Street, is that there is a superior

hypothesis available—the "adaptive link account"—according to which moral judgment evolved because it encouraged our ancestors to behave in an adaptive manner in the social sphere. She writes that the latter is to be preferred on grounds of parsimony (inter alia), because "the tracking account obviously posits something extra that the adaptive link account does not, namely independent evaluative truths" (ibid.: 129). As with Ruse's argument, however, one might complain that this appeal to parsimony has traction only if the "independent truths" in question involve positing ontological categories over and above those one antecedently accepts, and this is not obviously so for many forms of moral realism.

An evolutionary debunking argument with even weaker ambitions is one that I have myself developed (in Joyce 2006), which argues only for thesis 1. According to this view, the availability of a plausible genealogical theory of moral judgment—one which neither implies nor presupposes that these judgments are true—removes any justification which these judgments may have been accorded (on the grounds of epistemological conservatism, for example). Note that the conclusion of the argument is not that any moral judgment is *false* (contra Mason's interpretation (2010)),[6] and that the purported *objectivity* of morality plays no role in the argument at this point (contra Kahane's interpretation (2011)); if moral judgments were innate but *subjective* the same form of argument might apply.[7] But this evolutionary debunking argument does not go so far as to endorse thesis 2, thus leaving open the possibility that justification may be reinstated. This epistemological undermining is in a sense provisional: It represents a *challenge* that the supporter of moral belief must rise to overcome. One way of meeting the challenge would be to defend a form of moral naturalism according to which moral facts are identical to or supervene upon the items accepted in the evolutionary genealogy of morals. Therefore on an earlier occasion I supplemented my challenge with an attack on moral naturalism, centered on the contention that morality is essentially imbued with a kind of practical authority (conceptually speaking) which no combination of natural properties can provide. Thus while I have argued that empirical evolutionary discoveries are sufficient to create a substantive burden for the moralist, I also recognize the need to appeal to *a priori* metaethical methods in bolstering the challenge.

There are many more nuanced ways that evolutionary findings might influence moral philosophy. The foregoing discussion has confined itself to outlining two very broad programs—vindication and debunking—that purport to draw weighty ethical conclusions from biological discoveries, and identifying significant challenges for both.

---

[6] On other occasions I have certainly accepted something close to thesis 4 (see Joyce 2001), but did not purport to establish it on evolutionary grounds.

[7] For example, suppose we found ourselves believing that some actions are required because they are commanded by an all-powerful divine entity. The *non-objectivity* of these requirements (something that can be maintained because the requirements depend on some entity's mental states) would not prevent the deployment of a genealogical debunking argument. For example, any plausible adaptational hypothesis concerning why our ancestors profited from these beliefs would presumably not make reference to their successfully tracking the mental states of any such entity. Confirmation of such a hypothesis would thus account for our beliefs in a way that neither implies nor presupposes their truth; and this, it can be argued, has implications for the epistemological status of these beliefs.

References

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

Ayala F. 2010. "The difference of being human: Morality." *Proceedings of the National Academy of Sciences* 107: 9015-9022.

Boyd, R. & Richerson, P. J. 1985. *Culture and the Evolutionary Process*. University of Chicago Press.

Brown, S. 2008. *Moral Virtue and Nature*. New York: Continuum

Buller, D. 2006. *Adapting Minds*. Cambridge, MA.: MIT Press.

Buss, D., Haselton, M., Shackleford, T., Bleske, A., & Wakefield, J. 1998. "Adaptations, exaptations, and spandrels." *American Psychologist* 53: 533-548.

Campbell, R. 1996. "Can biology make ethics objective?" *Biology and Philosophy* 11: 21-31.

Casebeer, W. D. 2003. *Natural Ethical Facts*. Cambridge, MA.: MIT Press.

Chomsky, N. 1967. "Recent contributions to the theory of innate ideas." *Synthese* 17: 2-11.

Chomsky, N. 1987/1990. "On the nature, use, and acquisition of language." In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell. 627-646.

Confer, J., Easton, J., Fleischman, D., Goetz, C., Lewis, D., Perilloux, C., & Buss, D. 2010. "Evolutionary psychology: Controversies, questions, prospects, and limitations." *American Psychologist* 65: 110-126.

Conway, L. III & Schaler, M. 2002. "On the verifiability of evolutionary psychological theories: An analysis of the psychology of scientific persuasion." *Personality and Social Psychology Review* 6: 152-166.

Darwin, C. 1879/2004. *The Descent of Man*. London: Penguin Books.

Dennett, D. C. 1995. *Darwin's Dangerous Idea*. NY: Simon & Schuster.

de Waal, F. B. M. 1992. "The chimpanzee's sense of social regularity and its relation to the human sense of justice." In R. D. Masters & M. Gruter (eds.), *The Sense of Justice: Biological Foundations of Law*. Newbury Park, CA: Sage Publications. 241-55.

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

de Waal, F. B. M. 2006. *Primates and Philosophers*. Princeton University Press.

Dwyer, S. 2006. "How good is the linguistic analogy?" In P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind, Volume 2: Culture and Cognition*. Oxford: Oxford University Press. 237–55.

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

Fraser, B. 2010. "Adaptation, exaptation, by-products and spandrels in evolutionary explanations of morality." *Biological Theory* 5: 223-227.

Garner, R. 2010. "Abolishing morality." In R. Joyce & S. Kirchin (eds.), *A World Without Values*. Dordrecht: Springer Press. 217-233.

Gould, S. J. 1991. "Exaptation: A crucial tool for evolutionary psychology." *Journal of Social Issues* 47: 43-65.

Gould, S. J. 1992. "Male nipples and clitoral ripples." In his *Bully for Brontosaurus: Further Reflections in Natural History*. NY: W. W. Norton. 124-138.

Gould, S. J. & Lewontin, R. C. 1979. "The spandrels of San Marco and the Panglossion paradigm: A critique of the adaptationist programme." *Proceedings of the Royal Society: Biological Sciences, Series B* 205: 581-598.

Gould, S. J. & Vrba, E. 1982. "Exaptation: A missing term in the science of form." *Paleobiology* 8: 4-15.

Griffiths, P. 2002. "What is innateness?" *Monist* 85: 70-85.

Haidt, J. & Bjorkland, F. 2008. "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues." In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA.: MIT Press. 181-217.

Haidt, J. & Joseph, C. 2004. "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues." *Daedalus* 133: 55-66.

Hamlin, J. K., Wynn, K., & Bloom, P. 2007. "Social evaluation by preverbal infants." *Nature* 450: 557-559.

Hamlin, J. K., Wynn, K., & Bloom, P. 2010. "Three-month-old infants show a negativity bias in social evaluation." *Developmental Science* 13: 923-929.

Hauser, M. 2006. *Moral Minds*. NY: Harper Collins.

Hinckfuss, I. 1987. "The moral society: Its structure and effects." *Discussion Papers in Environmental Philosophy* 16. Canberra: Philosophy Program (RSSS), Australian National University.

Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

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

Joyce, R. 2012. "The many moral nativisms." In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (eds.), *Cooperation and its Evolution, Vol. 2: Agents and Mechanisms*. Cambridge, MA.: MIT Press.

Kahane, G. 2011. "Evolutionary debunking arguments." *Noûs* 45: 103-125.

Kant, I. 1783/1985. *Groundwork to the Metaphysics of Morals*. H. J. Paton (trans.). London: Hutchinson.

Ketelaar, T. & Ellis, B. 2000. "Are evolutionary explanations unfalsifiable? Evolutionary psychology and the Lakatosian philosophy of science." *Psychological Inquiry* 11: 1-21.

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

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

Laurence, S. & Margolis, E. 2001. "The poverty of the stimulus argument." *British Journal for the Philosophy of Science* 52: 217-276.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

Mameli, M. & Bateson, P. 2007. "The innate and the acquired: Useful clusters or a residual distinction from folk biology?" *Developmental Psychobiology* 49: 818-831.

Mason, K. 2010. "Debunking arguments and the genealogy of religion and morality." *Philosophy Compass* 5: 770-778.

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

Miller, G. 2000. *The Mating Mind*. Doubleday.

Miller, G. 2007. "Sexual selection for moral virtues." *Quarterly Review of Biology* 82: 97-121.

Nichols, S. 2005. "Innateness and moral psychology." In P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind: Structure and Contents*. New York: Oxford University Press. 353-430.

Nietzsche, F. 1887/1994. *The Genealogy of Morals*. C. Diethe (trans.); K. Ansell-Pearson (ed.). Cambridge: Cambridge University Press.

Noë, R. 2001. "Biological markets: Partner choice as the driving force behind the evolution of mutualisms." In R. Noë, J. van Hoof, & P. Hammerstein (eds.), *Economics in Nature: Social Dilemmas, Mate Choice, and Biological Markets*. Cambridge: Cambridge University Press. 93-118.

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

Prinz, J. 2008. "Is morality innate?" In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*. Cambridge, MA.: MIT Press. 367-406.

Prinz, J. 2009. "Against moral nativism." In D. Murphy & M. Bishop (eds.), *Stich and his Critics*. Malden, MA.: Wiley-Blackwell. 167-189.

Richards, R. J. 1986. "A defense of evolutionary ethics." *Biology and Philosophy* 1: 265-293.

Richerson, P. J. & Boyd, R. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago 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. 2006. "Is Darwinian metaethics possible (and if it is, is it well-taken)?" In G. Boniolo & G. de Anna (eds.), *Evolutionary Ethics and Contemporary Biology*. Cambridge: Cambridge University Press. 13-26.

Ruse, M. 2009. "Evolution and ethics: The sociobiological approach." In M. Ruse (ed.), *Philosophy After Darwin*. Princeton, NJ: Princeton University Press. 489-511.

Ruse, M. & Wilson, E. O. 1985. "The evolution of ethics." *New Scientist* 108: 50-52.

Smetana, J. G. 1993. "Understanding of social rules." In M. Bennett (ed.), *The Development of Social Cognition: The Child as Psychologist*. NY: Guildford Press: 111-141.

Smetana, J. G. & Braeges, J. L. 1990. "The development of toddlers' moral and conventional judgments." *Merrill-Palmer Quarterly* 36: 329-346.

Sober, E. 2000. "Psychological egoism." In H. LaFollette (ed.), *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell. 129-148. [##this volume??##]

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

Spencer, H. 1879/2011. *The Data of Ethics*. J. Turner (ed.). NJ: Transaction.

Spencer, H. 1851. *Social Statics*. London: J. Chapman. Sterelny, K. 2010. "Moral nativism: A sceptical response." *Mind & Language* 25: 279-297.

Stich, S. 2007. "Evolution, altruism and cognitive architecture: A critique of Sober and Wilson's argument for psychological altruism." *Biology and Philosophy* 22: 267-281.

Street, S. 2006. "A Darwinian dilemma for realist theories of value." *Philosophical Studies* 127: 109-166.

Tomkins J. L. & Brown, G. S. 2004. "Population density drives the local evolution of a threshold dimorphism." *Nature* 431: 1099-1103.

Turiel, E. 2002. *The Culture of Morality: Social Development, Context, and Conflict*. Cambridge: Cambridge University Press.

Weinstein, D. 1998. *Equal Freedom and Utility: Herbert Spencer's Liberal Utilitarianism*. Cambridge: Cambridge University Press.

Westermarck, E. 1906. *The Origin and Development of the Moral Ideas, Vol. 2*. London: MacMillan.

Wilson, E. O. 1978. *On Human Nature*. Cambridge, MA.: Harvard University Press.

Zahavi, A. 1977. "The cost of honesty: Further remarks on the handicap principle." *Journal of Theoretical Biology* 67: 603-605.