

## *The many moral nativisms*

Richard Joyce

Penultimate draft of the paper appearing in *Cooperation and its Evolution*, edited by K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (MIT Press, 2013).

### Introduction

John Stuart Mill’s opinion that “moral feelings are not innate, but acquired” (Mill 1861, 527) was, in the estimation of Charles Darwin, destined to be judged as “a most serious blemish” on that moral philosopher’s future reputation (Darwin 1879/2004, 121). But Darwin’s prophesy has so far proved incorrect; Mill’s opinion on the matter has hardly been commented upon, let alone decried. Indeed, the whole question of the origin of human morality received remarkably little discussion in the century or so after Darwin’s *Descent*.<sup>1</sup> The last decade, however, has seen the question placed back on the agenda. The emergence of *fin de siècle* Evolutionary Psychology—and in particular its pioneers’ decision to focus on the moralistic trait of “cheater detection” as their favorite case study (see Cosmides & Tooby 1992)—has prompted burgeoning debate about moral nativism. While this debate has yet to mature, and though one of its striking characteristics is a tendency for claims to be pressed (both for and against) with a confidence disproportionate to available evidence, we nevertheless might reasonably hope for genuine progress in the foreseeable future. Before that progress can occur, however, we need to understand the hypothesis. Currently there are a number of points of significant imprecision in the debate over moral nativism that often pass unnoticed and which lead to seemingly opposed factions speaking at crossed purposes. I think it is fair to say that we are at present in the same state which William Darwin (in a letter to his father) attributed to Mill: of being “rather in a muddle on the whole subject.”<sup>2</sup>

In previous works I have advocated moral nativism (Joyce 2006a, 2006b)—though I did so provisionally and cautiously; my objective was concerned more with clarification than all-out endorsement. Advocating moral nativism is not my intention in this chapter; my goal here is principally diagnostic. I will highlight three places where the nativist/non-nativist debate fragments in such a way that it ceases to be clear what the hypothesis is that is under dispute. In two of the three problematic places the options for reinstating precision are reasonably well defined, so my conclusion is that disputants simply need to take care to specify which understanding of the hypothesis is under discussion. In the case of the third, however, my attitude is rather more pessimistic. Here, it seems, we find at the heart of the debate an

---

<sup>1</sup> Of course, one would have little trouble assembling a list of books and articles from 1880 to 1980 (say) which would appear to counter this claim (Edvard Westermarck’s works in particular spring to mind); but I would maintain that this list—though superficially impressive if gathered in a footnote—still constitutes “remarkably little attention” for a century’s worth of intellectual labor on the topic.

<sup>2</sup> Darwin Archives: DAR88.76-77. Charles had evidently asked William to read and summarize Mill’s *Utilitarianism* for him while he (Charles) was preparing the second edition of *Descent*. Given that the point of this delegation of labor was to discern Mill’s views on the origin of the moral sense, I cannot resist remarking that it was William who had many years earlier been the subject of his father’s article “A biographical sketch of an infant,” and whose “first sign of moral sense” was observed at just over a year old (Darwin 1877: 291).

inchoate concept—that of *moral judgment*—regarding which the options for precisification are not well understood, and for which any stipulative specificity appears more of a misleading distortion than a welcome clarification. One possible consequence of this is that on some legitimate conceptions of *moral judgment* moral nativism is true, but on other equally legitimate conceptions moral nativism is false. And if there is no satisfactory way of deciding among these conceptions then the debate over moral nativism would be undecidable—not just in the sense that we lack decisive data, but in the sense that there is really no fact of the matter.

### **The first node of imprecision: Innateness**

In its crudest form moral nativism is the view that human morality is innate. What might be meant by “human morality” is a question that will occupy much of this chapter, but first our attention should pause on what is meant by “innate.” Some participants in the debate over moral nativism know what they mean by “innate,” but many employ an intuitive folk notion that doesn’t withstand critical scrutiny. Of those that do have a clear view of what they mean, not all mean the same thing.

The folk notion of innateness is a blend of several sub-clusters of ideas. One such group of ideas pertains to a trait’s being present at birth, to its being not learned, to its being determined by genes rather than environment, to its being developmentally robust in the face of environmental variation. Another idea central to innateness is the Darwinian notion of a trait’s existing because it was selected for by the process of natural selection—that is, of a trait’s being an adaptation. Another is the essentialist idea of a trait’s being species-typical: present in all members of the species or at least in all “normal” members. (For diagnosis and discussion of such options, see Griffiths 2002; Mameli & Bateson 2007; Mameli 2008.)

These ideas are not all equally scientifically respectable, and, more to the point, they are far from co-extensional. Down’s Syndrome is present at birth, genetically influenced, and developmentally robust, but is not an adaptation. The possession of a certain stone-knapping technique may satisfy the criteria for being an adaptation (it may be transmitted from parent to offspring and may owe its existence to the fact that it enhanced reproductive fitness), but is neither non-learned nor developmentally robust. And so forth. Hence we must reject the common unexamined presupposition that these phenomena more-or-less come together and thus can be treated as facets of a single “cluster” concept. In light of the way these disparate ideas get lumped together, Matteo Mameli disparagingly refers to *innateness* as a “clutter” concept (2008).

In the literature on moral nativism, two conceptions of innateness are most conspicuous: an evolutionary conception and a developmental conception. A typical statement of the evolutionary conception comes from Jesse Prinz, who sums up moral nativism as the claim that “morality is an evolved capacity” (2009, 168). I have myself described moral nativism as the view that “morality (under some specification) ... is to be explained by reference to a genotype having granted ancestors reproductive advantage” (Joyce 2006, 2). On this view, moral nativism is the claim that morality is a Darwinian adaptation.

Standing in contrast to this is the developmental conception, according to which the emergence of the trait is buffered against variation in the developmental environment (Ariew

1996, 1999). Stephen Stich and Chandra Sripada use such a conception when they write that “we can consider a normative rule to be innate if various genetic and developmental factors make it the case that the rule would emerge ... in a wide range of environmental conditions” (2006, 299).

These two conceptions of innateness are by no means co-extensional. The trait of morality might be a specific adaptation but may nonetheless require particular structured environmental inputs in order to become manifest. If such inputs were reliably available in the environment in which morality evolved, then there would be no selective pressure to make the developmental emergence of morality robust in the face of environmental perturbation. The reverse is also true: Human morality may be developmentally canalized while not being an adaptation. Some of the well-known ways by which traits may become canalized without being adaptations—genetic drift, mutation, genetic disease—are admittedly far-fetched in the case of morality. However, one way is entirely plausible: that morality is a by-product of other adaptations. (This possibility shall be discussed later.)

Clearly, this introduces potential confusion into the debate over moral nativism, for it allows that moral nativism may be true in one respect but false in another. Even when advocates of a particular view are conscientious in articulating which thesis they mean to defend or attack, casual readers may miss the qualification. 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 is an adaptation,” but adds that “even if that is right, it does not follow that this capacity is innate” (2010, 280). If such comments are overlooked, however, then one might gain the impression that Sterelny is in the same camp as other opponents of moral nativism when in fact these others are skeptical of the *adaptational* nativist thesis. More worryingly, one may gain the erroneous impression that Sterelny is in the opposing camp to someone like myself, who has advocated the adaptational nativist thesis, when it is entirely possible that we agree on everything of substance.

We have seen that in assessing the thesis of moral nativism possible misunderstandings lurk around the term “innate.” Yet the possible misunderstandings surrounding the term referring to the trait in question—“human morality”—are even greater. In subsequent sections I will tease this matter apart into two further particular points of imprecision, but first I will introduce the general problem via a discussion of altruism and Darwin’s views on moral nativism. My reason for doing so is as follows. I want to demonstrate that the trait in question, human morality, is difficult to define with any precision—that it admits of more liberal and more strict characterizations. In order to illustrate this, it is useful to begin with another trait, altruism, that is not a million miles from morality but which is pretty clearly not the same thing. Identifying the difference between altruism and morality forces us to ask what exactly the trait of “morality” is. Darwin’s own views are worth discussing here because he begins with prosocial attitudes, like altruism, which he then supplements with further psychological traits in order to achieve something which, he believes, deserves the label “the moral sense.” (Darwin, obviously, is focused on evolutionary rather than developmental emergence.) This transition from non-moral organism to moral organism is exactly what we are interested in. But Darwin’s efforts also exemplify the difficulty and obscurity of the

task—the fact that it is radically unclear what an adequate account of the transition from the non-moral to the moral would have to involve. There is, I wish ultimately to argue, no single answer to this question.

### **From altruism to Darwin**

It is standard to distinguish two forms of altruism: psychological and evolutionary. An action is psychologically altruistic if and only if it is motivated by an ultimate desire for the well-being of some other organism. A behavioral trait is evolutionarily altruistic if and only if it benefits another at some cost to the individual, where benefits and costs are understood in terms of reproductive fitness. (It must be added that the trait has been selected for *because* it benefits another, otherwise one ends up counting as altruistic such things as a sea turtle's drive to lay its eggs on the beach, which makes its hatchlings such easy prey for seagulls.) The former is an articulation of a vernacular notion, whereas the latter is very much a term of art.

The extensive literature ostensibly concerning the “evolution of altruism” often fudges this important distinction, and, indeed, frequently concerns neither. Consider the so-called altruistic behavior of bees. It is surely not psychologically altruistic (since bees simply lack the motivational prerequisites), but nor is it obviously evolutionarily altruistic: William Hamilton's breakthrough work on kin selection (1964) demonstrated how the individual bee who dies to save her nest-mates is in fact advancing her own inclusive fitness. Or consider the reciprocal grooming behavior of primates (see Schino and Filippo 2010). If the explanation of primate A's tendency to take the time and effort to groom primate B is that this increases the probability of A's being groomed in return, then in performing this behavior A is reproductively better off (eventually) than if it did not. (See West et al. 2007.) Similar considerations pertaining to hunting lions, mobbing birds, meerkats on sentry duty, etc., will also reveal neither psychological nor evolutionary altruism. For this reason, it is best to call such behaviors simply “cooperation” (leaving this an intuitive term), which then allows the questions of whether these cooperative behaviors are also instances of psychological altruism or evolutionary altruism to be substantive inquiries. (For discussion of how true evolutionary altruism is possible, see Sober 1988; Sober & Wilson 1998.)

Without pausing to investigate the details of how much cooperation in nature really is evolutionarily altruistic, one can at least safely say that cooperation often turns out to be evolutionarily selfish, in the sense that the cooperative behavior ultimately enhances the actor's reproductive fitness better than not cooperating. The temptation that it is crucial to resist is thinking that this evolutionary selfishness has any bearing on psychological selfishness. Organisms that do not have psychological states at all, such as plants, may be evolutionarily selfish or altruistic. In order to satisfy the prerequisites for being *psychologically* altruistic or selfish, a creature must be able to have ultimate motives concerning others' or their own welfare, which requires them to have the concepts of *other* and *self*. The only creatures for which we can be confident of the satisfaction of these prerequisites are humans.

I will take it as obvious that the mere fact that a behavioral trait is to be explained by reference to *evolutionary* altruism is insufficient to make the introduction of talk of

“morality” appropriate. A plant may have evolutionarily altruistic traits, but the plant neither makes moral judgments nor is a suitable subject of our moral appraisals. But it is not so obvious that there is no connection between *psychological* altruism and morality, so this requires some discussion. It is particularly important here because a plausible case can be made that psychological altruism in humans is innate, hence this may have direct implications for the prospects of moral nativism.

The details of the argument for nativism concerning psychological altruism need not delay us; a sketch will suffice for present purposes. The argument concerns *evolutionary* nativism rather than developmental nativism, and has been advocated by Elliott Sober (2000). Sober’s principal opponent is the psychological egoist, who holds that all human actions are performed with the ultimate motive of benefiting the actor. Given that natural selection has clearly forged humans to be cooperative in certain ways—at the very least, caring for our offspring—Sober wonders what kinds of psychological mechanisms would be likely to be favored to govern these cooperative tendencies. Assuming that it is adaptive to come to the aid of one’s children when they are in distress, for example, what is the better psychological set-up? On the one hand, we can envisage a parent motivated to provide aid simply because he loves his daughter—he cares directly for her in such a way that a perceived threat to her welfare directly prompts action. On the other hand we can imagine the egoistic parent: moved via a combination of the belief that his daughter’s suffering has a negative effect on his own welfare plus his love for himself. One might plausibly claim that the former mechanism is more reliable and less complicated—and thus, *ceteris paribus*, more adaptive—than the latter. By analogy, a person prompted to withdraw her fingers from a flame *by pain* seems moved by a more reliable and less complicated process than a person who forms a belief about the bodily damage caused by fire and calculates the costs and benefits of action versus non-action. This argument may not be without problems (see Stich 2007), but here my intention is not to evaluate or endorse the argument, but rather to examine what would follow—or, more precisely, what would *not* follow—if it were sound.

We have seen that in order to be psychologically altruistic a creature needs to be fairly cognitively sophisticated, but it doesn’t follow that the creature is therefore capable of making moral judgments. This truism is potentially muddied by the fact that the only clear-cut case of a species capable of psychological altruism (and selfishness) is also the only clear-cut case of a species of which we speak in moral terms: namely, humans. Still, the conceptual distinction does not seem difficult to discern. One can imagine members of a cognitively sophisticated social species, motivated by love and altruistic tendencies towards their fellows, but who fail to “moralize” these feelings—who are, in fact, constitutionally incapable of making a moral judgment. Such creatures have powerful desires to see their loved ones flourish, but cannot conceive of actions satisfying those desires as morally right or obligatory.

It might be conceded that these imaginary creatures don’t make moral judgments but maintained that they are at least morally *praiseworthy* (that is, that they warrant *our* moral judgment). But upon reflection even this is unclear. After all, altruistic motives can prompt someone to act in a morally despicable manner. Consider a mother who genuinely adores her child, and who poisons all the other children at the sports day so her child can win. In any case, it seems misguided to identify moral nativism with the claim that the trait of *being morally praiseworthy* is innate. Such a proposal would lead straight into a meta-ethical

quagmire from which the debate is unlikely ever to emerge. We are not primarily interested in the question of at what point, either in evolution or development, humans become *morally admirable*; we are interested in at what point they become capable of *making moral judgments*. Popular discussions of moral nativism with headlines like “Are we born to be good?” or “The moral animal” or “Chimps display morality” (etc.) blur this basic distinction, and in doing so spread more misunderstanding than illumination.

Once we focus nativism on the question of *moral judgment*, it becomes clear that we are asking about something different from (or perhaps more than) psychological altruism. None of this is to deny that the emergence of psychological altruism (both evolutionarily and developmentally) might be a crucial precursor to moral judgment; I’m not claiming that someone with an interest in moral nativism should dismiss all discussion of the emergence of psychological altruism as irrelevant. My claim is simply that moral judgment is not the same thing as altruism, and that establishing nativism about altruism does not establish moral nativism. Though this much seems assured, the natural further question of what exactly *is* required for moral judgment is much harder to answer.

As a way of illustrating this problem I turn now to Darwin’s views on the matter. Darwin undertakes the task of supplementing prosocial emotions (like altruism) with further psychological capacities in an attempt to “build” a human moral sense. Drawing attention to the difficulties inherent in this project is one of the goals of this chapter, so sketching his attempt is a useful exercise.

Darwin is no psychological egoist. He writes:

With respect to the impulse which leads certain animals to associate together, and to aid one another in many ways, we may infer that in most cases they are impelled by the same sense of satisfaction or pleasure which they experience in performing other instinctive actions. ... In many instances, however, it is probable that instincts are persistently followed from the mere force of inheritance, without the stimulus of either pleasure or pain. ... Hence the common assumption that men must be impelled to every action by experiencing some pleasure or pain may be erroneous. (1879/2004, 128)

He speaks frequently of the “social instincts” of animals—which include such affections as sympathy, love, and pleasure in the company of one’s fellows—and there is no doubt that Darwin considers these capacities to be psychological adaptations. But he is equally adamant that they do not suffice for a moral sense: “I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important.” He goes on:

[A]ny 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)<sup>3</sup>

---

<sup>3</sup> Darwin uses “moral sense” and “conscience” seemingly interchangeably. One interesting implication is that he sees the moral sense primarily in terms of *self-directed* moral evaluations—for that is what a conscience is. It seems to me, moreover, that this gives license to assume that when Darwin talks of a “moral sense” it is a faculty of making moral *judgments* that is under discussion. While I am aware that there is some room for debate about this assumption, here I’m willing to forego argument and treat it as a simplifying supposition.

What are these “intellectual powers”? First of all, Darwin thinks, one needs a good memory, in order to recall those times in the past when one has failed to act cooperatively and (as a result of one’s social instincts) felt dissatisfaction. One needs to recall that the benefits gained from failing to cooperate (i.e., the profits of defection) were fleeting. Second, the emergence of language allows that “the wishes of the community could be expressed, [and] the common opinion how each member ought to act for the public good, would naturally become ... the guide to action” (1879/2004, 122). Lastly, one needs the capacity to form *habits* of acting for the good of one’s fellows.

This might be interpreted as an argument for moral nativism (of the adaptational variety), but on another interpretation Darwin thinks of the moral sense as a kind of “spandrel” derived from faculties that evolved for other purposes.<sup>4</sup> In fact, he is explicitly undecided on the matter. Referring just to the social instincts, he writes 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). We will return to this distinction later; currently what interests me is how, precisely, the moral sense is supposed to emerge from these elements. My strategy will be to grant Darwin these ingredients and attempt to motivate doubt that we have enough to warrant the label “a moral sense.”

Darwin certainly has plenty of persuasive things to say about the evolution of the social instincts; on this topic he is squarely in his “comfort zone.” But his explanation of how certain “intellectual powers” get married to those instincts, resulting in a moral sense, is considerably sketchier and less convincing. The latter two ingredients listed in his initial presentation—language and habit—hardly get a further mention. (It is, besides, unclear whether he thinks of these two traits as *necessary* for a moral sense.) It is the role of *memory* that he mentions repeatedly and evidently judges of paramount importance. But the case is under-described at best.

Consider a creature brimming with altruistic sentiment for its conspecifics. I argued above that this alone does not suffice for a moral judgment. The creature doesn’t think that it *ought* to help its fellows; it doesn’t think of failure to help as *prohibited*; it doesn’t think that such failures warrant punishment or disapproval, or that helping merits praise. It simply *wants* to help. Yet suppose that occasionally the creature experiences temptations to do otherwise, since there are other competing instincts operative in its psyche. When this creature succumbs to such temptations, it enjoys the satisfaction of the tempting outcome (whatever it may be) and yet also feels bad because of the frustration of its natural desire to cooperate. Let us stipulate that the creature’s instincts are such that the pleasures achieved at the expense of cooperation tend to be short-lived. Let us now grant it the intellectual powers both to realize and to remember this fact. Thus, when temptation arises, the creature is able to deliberate along the lines of: “Well, that sure looks enticing, but I remember how rotten I felt last time I succumbed to temptation, so I’ll cooperate.” So now we have a creature with self-control in favor of cooperative behavior.

But where does the *moral judgment* emerge in this process? Acting cooperatively is still, essentially, just something that this creature *wants* to do. Compare a monkey that is often tempted to climb its favorite tree using the dangling outer branches, but who, through trial

---

<sup>4</sup> In previous work I have interpreted Darwin as a moral nativist; I now think that this is not straightforward.

and error, comes to learn that it is safer to ascend by the trunk. When faced with the temptation to dart up the dangly branches the monkey may pause and recall the bruises of earlier decisions. So now we have a creature with self-control in favor of climbing a tree via the trunk. But do we credit the monkey with anything like the judgment that climbing the outer branches is a *transgression*? If it does climb by the outer branches without mishap, we can imagine it thinking “That was a bit stupid, but, phew, I got away with it!” Where would be the *guilt*? Where would be the thought that it *deserves punishment* for its crime? Why would it take an interest in punishing other monkeys that exhibit foolish climbing habits?

In the case of the first creature whose instincts are in favor of cooperation, we need also to factor in the reactions of its conspecifics, but I don’t believe that this alleviates the puzzle. The conspecifics *don’t like it* when the individual defects on some cooperative enterprise, and we can imagine that their disappointment and anger is something that our individual will take into account. It controls itself by remembering how bad its failures to cooperate made it feel in the past, and when those failures are accompanied by its fellows expressing their anger with (say) violence and ostracization, then self-control will be all the easier since the negative repercussions of such failures will be even worse. Thus the influence of the conspecifics will certainly significantly strengthen the process of self-control, but it in no obvious way brings about a change in *kind* in the sorts of judgments and attitudes which we attribute to the individual.

It appears, therefore, that one can identify elements that seem important to moral judgment—such as the ideas of transgression, guilt, and desert—for which Darwin’s hypothesis does not account. In assessing this matter one needs to be wary of projecting one’s own “moralizing” thoughts onto the imaginary characters involved. It is difficult to cleanly imagine someone simply *not wanting* to perform noncooperative actions (in part because she recognizes that other parties *don’t want* her to) without positing the seemingly innocuous extra assumption that she also judges that she *ought not* perform those actions. It is natural for us to assume that as our imaginary creature forms the habit of acting cooperatively, surely at some point it “internalizes the norm”: its *expectation* of negative outcomes morphs naturally into the thought that such outcomes are *warranted*; its *desire* for its fellows’ welfare gradually begets the judgment that acting for their welfare is *desirable*; it moves from habitually not wanting something to judging it prohibited; and so forth. But assuming that this transition occurs naturally is exactly what we must not do in this context, for how such a transition occurs is precisely what is under scrutiny.

Darwin brings the discussion to the edge of “moralization” but it is not obvious that he succeeds in crossing the conceptual gap. Perhaps the ingredients he provides suffice for a thin notion of moral judgment, but there is a richer folk conception whose evolutionary emergence remains mysterious. As we shall see, the same can be said of some modern participants in this debate: They provide ingredients that may be adequate to account for moral judgment in some attenuated sense but which fail to explain important components of a robust conception of moral judgment. Thus the debate founders not merely through lack of empirical data, but through an absence any single phenomenon uniquely deserving of the name “moral judgment.” Before discussing this matter further, however, I should like to note another source of confusion about the nature of the trait whose origin is under discussion.

## The second node of imprecision: Content vs. concept

It is important to distinguish between moral concepts and moral judgments. Let us say that a complete paradigm moral judgment consists of the application of a moral concept, like *moral wrongness*, to a general subject, like *incest*, or to a particular subject, like *John and Mary's incestuous relation*.<sup>5</sup> Given this framework, we can identify another way in which moral nativist hypotheses may vary.

One version of moral nativism will allow that certain complete moral judgments are innate. There is certainly nothing to be said in favor of the claim that complete moral judgments concerning *particulars* are innate. For example, to hold that the judgment "John and Mary's incestuous relation is morally wrong" is an adaptation would involve accepting that our ancestors somehow knew about the individuals John and Mary, formed a moral opinion about what they got up to, and that this opinion enhanced reproductive fitness. Given that a great many of our moral judgments do concern particulars, nativism about complete judgments is going to be utterly implausible for a great many of our moral judgments. Even for those moral judgments that take universals as subjects, nativism concerning the complete moral judgment is feasible only when the subject is something that was present in the environment of evolutionary adaptiveness (the EEA). One may, for example, countenance nativism for "Incest is wrong," but nativism for "Shoplifting is wrong" is a non-starter.

Another version of nativism eschews any commitment to complete judgments being innate, and prefers the image of a moral faculty as a "toolkit" of moral *concepts*, with the individual's socialization process as the sole determinant of to which subjects these concepts get attached. Thus according to this hypothesis (expressed in simplistic terms), a concept like *moral wrongness* is innate, and one social environment may lead the individual to apply the concept to *incest*, another environment may lead the child to apply it to *John and Mary's incestuous relationship* but not to *Ptolemy and Cleopatra's incestuous relationship*, while yet another may lead the child not to apply the concept to any incestuous relationship.

These two nativist positions represent extremes, between which lie a variety of hypotheses. Some allow that a few broad abstract moral principles are innate but that the environment sets the parameters of how these create specific moral judgments (Hauser 2006).

---

<sup>5</sup> This statement may seem metaethically question-begging and also surprising in light of other claims I have just made, so a couple of quick explanations are called for. First, at this stage of the discussion I don't intend this notion of "applying a concept" to be theoretically deep, thus my claim is meant to be metaethically neutral. I take it that the locution "applying a concept" is something which even the modern noncognitivist will seek to accommodate. Simon Blackburn's quasi-realist program sets out to "earn the right" to such realist-sounding talk but from an anti-realist position that eschews any genuine metaphysical commitment to such entities. (See Blackburn 1993, 1998.) Second, given the emerging worries about the indeterminacy surrounding the notion of *moral judgment*, one may wonder on what grounds I can confidently make such an assertion. The answer is that even if there are thinner and richer explications available of the notion of *moral judgment*—such that the former counts certain things as moral judgments that the latter will not—nevertheless there is surely a class of paradigm instances of moral judgments to which all parties will agree. Of these paradigms, though there may remain disagreement concerning in virtue of what they count as moral judgments, it hardly follows that we can say nothing about their characteristics. The statement to which this note is appended is intended to be just such a platitudinous description.

Some allow that content is learned but that the moral sense comes “prepared” to latch on to certain domains more easily than others (see Haidt & Joseph 2004; Sripada 2008).<sup>6</sup>

Even with some options in the moral nativism spectrum sketched in so heavy-handed a manner, we have seen enough to recognize that evidence favoring one version of moral nativism will not favor another. As a way of illustrating the muddle that ensues, I will examine the debate over moral universals.

In fact, even if all parties were in complete concurrence regarding which trait is under scrutiny, the place of universals in the debate over nativism would be far from straightforward. The tempting assumption that if a trait is innate then we can expect to find it manifest everywhere must be rejected. If one is focused on developmental innateness, then many innate traits are not universal (Down’s syndrome, eye color, lactose tolerance, etc.). If one is discussing adaptational innateness, then innate traits may well require substantial environmental input—input that may have been reliably present in the EEA but is absent, patchy, or distorted in the modern environment. I intend to put these important complications aside, however, in order to focus on another simpler point about universals. For the sake of argument let us allow the assumption that innate traits will reliably emerge and thus tend towards universality. The question is: For what kind of universals should we be looking? And the answer is: It depends which version of moral nativism is under scrutiny.

In one of a series of papers arguing against moral nativism, Prinz discusses three possible moral universals: *don’t harm innocent people*; *respect and obey authorities*; and *incest is prohibited* (Prinz 2009; see also Prinz 2008a, 2008b, forthcoming). He carefully examines historical and anthropological evidence in an attempt to find counter-examples to the claim of universality for each, thus discrediting moral nativism. But the limitations of this strategy should by now be clear: Many moral nativisms will not hold that such complete moral judgments are innate.<sup>7</sup>

This is not to say that Prinz’s efforts are wasted. Certain versions of moral nativism may well claim that precisely these three complete moral judgments are innate, and I share Prinz’s determination to reject such views. Prinz, moreover, knows that he is bothering only one form of moral nativism. He is aware of the kind of “toolkit” moral nativism mentioned earlier—which holds no complete moral judgment to be innate but rather postulates innate moral concepts. Prinz labels this kind of moral nativism “minimal” (2009) and “weak” (forthcoming). I confess to finding this labeling system unfortunate, since it allows the anti-nativist to proceed by first refuting the “strong” versions of moral nativism (the kinds that were never terribly plausible in the first place), thus giving the impression of the moral nativist retreating to an ever weaker position in a desperate bid to defend his/her hypothesis. The rhetorical narrative this suggests is inaccurate and is exasperating to anyone who begins

---

<sup>6</sup> Note my avoidance of speaking of “innate moral *knowledge*”—an unnecessary practice that seems to beg several large questions. Moral nativists who seemingly lacks such qualms include Sue Dwyer (Dwyer 2009; Dwyer et al. 2010) and John Mikhail (2008).

<sup>7</sup> Another potential problem is that Prinz sets out to investigate the existence of *cultural* universals, whereas if nativism did imply universality, we should be examining evidence of *psychological* (i.e., individualistic) universals. For the sake of argument I’ll play along with the focus on cultural universals. See Buller 2006, 457ff. for critical discussion.

with a desire to defend “toolkit” moral nativism while agreeing wholeheartedly that there is little to be said in favor of the more content-complete versions of nativism.

If we are investigating evidence for and against universality, and have the more plausible “toolkit” kind of moral nativism in mind, then we should be examining whether there are any cultures that lack moral judgments altogether. If one culture thinks that incest is morally acceptable while another judges it repugnant, this is no counter-example to universality, for both cultures are still evaluating the world in moral terms. All too often the debate has revolved around the question of whether there exist “moral universals,” but if I am correct then this is misguided; what we should be investigating is whether *having a system of moral judgments* is a human universal. And while it is not my intention on this occasion to press the case in favor of this latter hypothesis, it is reasonable to suppose that the prospects of its being true are far better than the likelihood of finding moral universals.

Prinz is certainly unable to provide a counterexample. At one point he mentions the Ik group of Uganda, famously described by anthropologist Colin Turnbull (1972) as a “vicious people” with “sadistic customs.” We now know that Turnbull’s account of the Ik was flawed in numerous ways (see Heine 1985; Knight 1994), but even if that were not so, the “viciousness” of which he spoke is compatible with the Ik having a moral system—one that might seem blighted and alien to us, but a moral system nonetheless. Indeed, when, several years later, the Ik elders heard of how Turnbull had portrayed them to the world, they were angry that he had “spoilt” their reputation, and threatened to make him “eat his own faeces” if he ever showed his face again (Heine 1985, 3). To the extent that they thought that Turnbull *deserved* this unenviable fate, the Ik proved themselves capable of wielding a moral concept.

Prinz doesn’t seriously think that the Ik lack any moral system. When he squarely addresses the “toolkit” version of moral nativism, he admits “I certainly don’t know of any exceptions to this claim” (2008a, 386). This concession forces a change of tactic in his pursuit of the non-nativist agenda: He moves from trying to provide counterexamples to universality and instead sets out to demonstrate that an appeal to nativism is not required to explain moral judgment; he endeavors to provide an empiricist explanation of the (possibly universal) phenomenon. In doing so he aims to discredit a focal argument in favor of moral nativism: the poverty of the stimulus (POS) argument. According to this argument, the capacities evident in moral cognition are acquired in a manner that far outstrips the information that is available in the learning environment. 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.<sup>8</sup> It is not my intention here to evaluate the prospects of a moral POS argument, but rather point out how progress gets confounded by distinct theoretical options being conflated.

One obvious way of countering a POS argument is to show that the stimulus is in fact a great deal less impoverished than one might have thought. Thus moral non-nativists are eager to point out how rich is the moral learning environment of the child. Shaun Nichols reminds us that “the child is exposed to *lots* of admonitions and instruction in the normative domain.

---

<sup>8</sup> For powerful criticism of this orthodoxy, see Cowie 1999.

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)

This is all undeniable. The child's moral world is richly structured, and the explicit moral instruction is coordinated and unrelenting.

It is not sufficient, however, simply to remark upon the wealth of the moral stimulus in a general way. We need to decide which version of moral nativism is under discussion, for this determines what kind of moral task it is whose acquisition process is under scrutiny. If our interest is in "toolkit" moral nativism, then focusing on how children acquire complete moral judgments is misleading; rather, our attention should be on how children acquire their basic moral conceptual tools. If this is the target trait, then wondering how children acquire the belief that shoplifting is wrong (say) would be a distraction (for I'm sure all parties can agree that they are taught it by adults); instead we should be wondering about how children acquire the concept of *moral wrongness* in the first place. Is the environment rich enough to provide them with *that*?

This is a crucial disambiguation to make before assessing the prospects of any moral POS argument, yet it still leaves progress hampered by a serious conceptual imprecision, for one is still left wondering "What is a moral judgment?" The possibility remains that moral nativism may be more plausible with certain conceptions than others. This is discernible in anti-nativist attempts to oppose the moral POS argument, as the following short review will demonstrate.

### **Some anti-nativist hypotheses**

The opponent of moral nativism will usually try to account for the human trait of making moral judgments by calling attention to other psychological traits that evolved or develop for other purposes. Often moral judgment is described as a byproduct or "spandrel" of these other traits. I will sketch a few anti-nativist views in order to give a flavor of the approach.

Prinz attempts to account for the evolutionary emergence of moral judgment from a cluster of other evolved faculties, each of which has a more general role. At the center of his argument is the view that moral judgments are emotional responses.<sup>9</sup> In one paper (2009), he proposes to construct a moral response out of emotions that are not distinctively moral: anger and sadness. We feel sad in many circumstances, but when we feel sad at having transgressed

---

<sup>9</sup> Or so Prinz claims when he's summarizing his view, but the more detailed presentation is rather more complicated. First, it turns out that having emotions is just the "standard" way to assess things morally (2007, 42). Second, moral judgments are linked by Prinz not directly to emotions but to *sentiments*—where a sentiment is a disposition to have an emotion (2007, 84). Thus Prinz has at least two "escape routes" should evidence come forward of moral judgments made with no emotional arousal. For further criticism of Prinz's view, see Joyce 2009.

against a norm, Prinz argues, then the sadness is called “guilt.” “Guilt is an accidental byproduct of sadness” (2009, #). In other works, Prinz develops a somewhat different empiricist hypothesis. In his 2008a he mentions not only the non-moral emotions, but some additional traits: *meta-emotions* (emotions directed at our own emotions or at others’ emotions), *perspective taking* (allowing for third-party concern), and other *non-moral preferences* (e.g., the “social instincts” which were Darwin’s starting point). The important point is that these are all general cognitive skills; thus if moral judgment is a natural byproduct of these traits the moral nativist would be defeated.

The view that emotion has a central role in moral judgment is also at the heart of Nichols’ attempt to provide an empiricist account of the origin of moral judgment (2005). Nichols allows that “rule nativism” might be reasonable, where the rules in question are non-hypothetical. “There is no obvious story about how the empiricist learner might come to acknowledge *nonhypothetical imperatives*” (2005, 357). He correctly argues that morality is but a proper subset of non-hypothetical rule systems, citing etiquette and institutional rules (e.g., of a gentlemen’s club) as involving non-moral but non-hypothetical imperatives (following Philippa Foot 1972). A key question, then, is what is distinctive about *moral* non-hypothetical imperatives. Nichols’ answer starts by noting the distinctive subject matter of morality—namely, that it pertains to harm.<sup>10</sup> Given this characterization of morality, the second ingredient in Nichols’ hypothesis is an innate affective mechanism that responds to suffering in others. This emotional response imbues a certain subset of non-hypothetical imperatives with a particular flavor (call it “moral”), picking them out as salient, resonant, and memorable. Nichols concludes:

[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)

Another anti-nativist argument comes from Sterelny, though, as noted earlier, he is focused more on the developmental trajectory than the evolutionary emergence of the trait. Like both Prinz and Nichols, Sterelny holds that one of the key psychological ingredients in a non-nativist explanation of moral judgment is emotion. He appears willing to endorse nativist hypotheses for emotional contagion, for sensitivity to interactions involving harm, and for the emotions associated with “reciprocation, sympathy, empathy, disgust, and esteem” (Sterelny 2010, 293). He argues at length that moral learning is largely a matter of generalizing from exemplars—which explains why moral intuitions can be fast and automatic—and also stresses that this would not mark moral learning as unusual (i.e., the faculties involved in prototype-comparison learning are general mechanisms). Sterelny further persuasively emphasizes the extremely rich and structured nature of the moral learning environment, arguing that the “parental generation engineers the informational environment in which the next generation develops, thus guaranteeing the development of moral competence” (294).

---

<sup>10</sup> Nichols is aware of moral norms that have nothing obvious to do with harm (concerning, e.g., cleaning the toilet with the national flag), but he states that “it is plausible that judgments about harm-based violations constitute an important core of moral judgment” (Nichols 2004, 7).

Sterelny concludes that moral norms “are grafted on top of our dispositions to respond emotionally” (292), that moral cognition “is a natural development of our existing emotional, intellectual and social repertoire” (293), and that moral cognition “develops from an interaction between emotions, exemplar-guided intuitions and explicit principles” (293).

Clearly, it is beyond the ambitions of this chapter to attempt to analyze or refute these proposals in detail; I aim to make a more general point. First I will pursue the same strategy as was deployed earlier against Darwin: taking the ingredients offered and questioning whether they suffice for making a moral judgment. My ultimate goal, however, is not to declare that all such arguments simply fail, but rather that there are different conceptions of moral judgment in play.

Consider, first, Prinz’s argument that guilt is just sadness directed at having transgressed against a norm. There appear to be important components of full-blooded guilt that remain unaccounted for. Sadness predicts social withdrawal, whereas guilt (unlike shame) urges reparative action (Tangney & Fischer 1995; Tangney, this volume). Extreme sadness cripples a person’s capacity to engage in everyday activities, whereas guilt, even acute guilt, is a burden that a person can usually shoulder while getting on with things. Even the manifestation of weeping that we associate with sadness we do not associate so readily with guilt (which is not to deny that guilt can cause a person to cry<sup>11</sup>). Indeed, language itself should be a giveaway here. We do have words for some special instances of sadness defined according to their object. “Grief,” for example, denotes sadness directed at the loss of someone or something dear to us. Notice that just as we can say “I feel grief about Fred’s death,” we can say “I feel really sad about Fred’s death,” and no one will bat an eye-lid. But compare the huge difference between saying “I feel guilty about having committed that crime” and “I feel sad about having committed that crime.”

Consider, second, Prinz’s argument that attempts to build moral judgment out of non-moral emotions (e.g., blame, which includes “other-directed emotions, such as anger, contempt, disgust, resentment, and indignation” (Prinz 2008a, 368-9)) combined with meta-emotions, third-party concern, and abstract ideas.<sup>12</sup> As a way of testing the adequacy of this empiricist hypothesis, let us imagine someone who satisfies all these components for one of the other-directed emotions that Prinz mentions: disgust. Suppose Ernie sees Bert vomit and feels disgust. Perhaps Ernie feels embarrassed at this response, or perhaps he is pleased with it; in either case he manifests meta-emotions. When Ernie thinks about some distant other person vomiting, he finds this idea pretty disgusting too; hence the emotion can be directed at third parties. Ernie is also capable of forming abstract ideas, so even the thought of vomit in some abstract sense makes him feel queasy.

It is clear that Ernie is pretty unhappy about Bert’s vomiting, but it is considerably less clear that he has made a full-blooded moral judgment about it. We apparently need not credit him with the ideas that vomiting is wrong, that Bert has transgressed, or that vomiters deserve reprimand (or that non-vomiters deserve praise). These, it will be noticed, are distinctly

---

<sup>11</sup> Yet when one pictures guilt prompting tears, it is natural to picture the scene as one where the transgressor is confronted and accused. By contrast, we have no trouble imagining the tears of sadness falling in private.

<sup>12</sup> Prinz adds the capacity for abstraction to his list of general mechanisms that account for moral judgment in Prinz forthcoming.

*cognitive* elements that are lacking in Prinz's account. If our conception of moral judgment privileges such cognitive elements, then Prinz's project must be deemed inadequate.

According to Nichols, core moral judgments concern harm prohibitions which are lent resonance and prominence by an innate affective program. One might also want to insist that a key element of moral norms (as opposed to other kinds of non-hypothetical norms) is that they have a special kind of practical authority. Foot, for example, discusses the Kantian idea that to transgress against a moral imperative is irrational, whereas transgressions against etiquette need not be. Elsewhere, I have followed John Mackie (1977) in suggesting that moral imperatives are conceptually "non-institutional" whereas those of etiquette are not (see Joyce 2001, 2011). Nichols doesn't deny this extra authority with which morality is imbued, but he argues that it comes into the picture later: as a consequence of the affective resonance of this class of norms. He writes that "the affective response seems to play a major role in determining the strength of one's normative commitments. ... [T]he affect-backed norms are treated as having justifications that go beyond the conventional" (2004, #).

But the nature of this connection remains puzzling to me. It can be granted that emotionally charged norms may be more memorable and seem more important. Yet it does not obviously follow that such resonant norms must also be accorded a stronger binding quality, that they will seem to hold independently of any institutional backing, that they will appear to require no further justification, or that one will be tempted to treat their violation as a form of irrationality. If affectively-underwritten norms happen to produce this air of practical authority, then this is a phenomenon requiring explanation. Until such an explanation is offered, then to the extent that one's conception of a moral judgment makes central this idea of special practical *authority*, Nichols' empiricist hypothesis doesn't pass muster.

The ingredients offered by Sterelny suffice for a social creature who is sensitive to harm situations, who feels empathy for his fellows, who generalizes from exemplars, for whom departures from the cooperative order are memorable and salient, and who, as a consequence, operates extremely well in his social world. But where is the morality? The language Sterelny uses does seem to acknowledge that there is at least some important element of morality that is *more* than the joint exercise of these capacities, for he writes of moral norms "developing from" and being "grafted on top of" these capacities. This seems correct, for it appears no great feat of the imagination to envisage a social creature who enjoys the traits allowed by Sterelny but who is nevertheless constitutionally incapable of making moral judgments concerning an action's meriting punishment, a norm's having convention-transcending practical authority, or even an outcome's being desirable (as opposed to being desired). It is, in other words, not hard to imagine a creature who enjoys all Sterelny's ingredients but for whom full-blooded moral cognition does not simply "develop." Hence, if one's conception of moral judgment privileges such cognitive accomplishments, then what is required is an explanation for why and how it does develop from these ingredients in the normal human case.

### The third node of imprecision: Moral judgment

From this review of some anti-nativist hypotheses a pattern has emerged. Anti-nativists tend to understand moral judgment in term of emotional traits which, they think, have more general psychological roles and thus are unlikely to count as mechanisms dedicated to the production of moral judgment. However, the ingredients they offer appear to leave certain more cognitive elements of moral judgment unaccounted for. Though I am tempted by the hard-nosed response of insisting that these cognitive components are *essential* to moral judgment and thus that these anti-nativist arguments fail, my considered stance is more pluralistic.<sup>13</sup>

I suggest that the notion of *moral judgment* is sufficiently pliable as to allow of different legitimate precisifications. A less demanding conception can be built largely out of emotional resources. To the extent that the less demanding conception might feel unsatisfying, in that it leaves certain cognitive elements of moral judgment unaccounted for, we must recognize the existence of a more demanding conception.<sup>14</sup> It is not a matter of there being two or more concepts; it's a matter of there being competing precisifications of the same somewhat indeterminate concept. A liberal conception will count as moral judgments items that the strict conception will not. And even for a paradigm moral judgment about which there is no doubt, the competing conceptions will disagree as to the criteria in virtue of which the item counts as a moral judgment. It's not a matter of our not knowing which is the correct conception (because we lack data); it's that there is no unique fact of the matter.

A similar view has been expressed in the useful comparison case of the human language faculty. Marc Hauser, Noam Chomsky, and W. Tecumseh Fitch—recognizing that “the word ‘language’ has highly divergent meanings in different contexts and disciplines” (Hauser et al. 2002, 1570)—distinguish between a faculty of language in a broad sense and in a narrow sense. The former, they hypothesize, consists largely if not entirely of capacities that humans share with other animals, while the latter (which is basically the capacity for linguistic recursion) is a uniquely human trait.

---

<sup>13</sup> In the past I have offered a fairly detailed description of what I take moral judgments to be, involving strong cognitive elements (Joyce 2006, chapter 2). This characterization has been criticized as being non-mandatory (see Machery & Mallon 2010), and, indeed, Stich finds it necessary to speak of “Joyce-style moral judgments” (2008, 234).

<sup>14</sup> I should point out that in the interests both of simple expression and playing along with an entrenched dialectic, I am drawing a line between “emotions” and “cognitions” in the orthodox ham-fisted manner. Of course the real distinction is nuanced and complicated. I should also say something to clarify the relation (or lack thereof) between the view under discussion and the literature on the neuroscience of moral judgment, in which the question of emotions versus cognitions looms large. Joshua Greene argues that some moral judgments (deontological ones) stem from emotional arousal while others (consequentialist judgments) flow from rational faculties. (See Greene et al. 2001.) Be that as it may, the deontological judgments that are prompted by emotional responses still, in my book, involve obvious cognitive elements. For example, judging that someone has an inalienable right to something (for which consequentialist considerations are irrelevant) involves the deployment of the hefty abstract concept *inalienable right*. Similarly, Jon Haidt's work (2001) may show that moral judgments are little more than post hoc rationalizations of knee jerk emotional responses, but this should not be confused with the claim that moral judgments are nothing more than emotional responses. Though Greene and Haidt (and others) underline the central role of emotion in moral judgment, they need not be interpreted as proponents of a less demanding conception of moral judgment.

But whereas Hauser, Chomsky, and Fitch do an admirable job of delineating the various skills and capacities involved in the two senses of “language faculty,” I feel somewhat pessimistic that the same can be done for “moral faculty,” for here, it seems to me, matters are considerably more nebulous. The three examples of non-nativists described above—Prinz, Nichols, and Sterelny—hardly present a univocal picture of what a liberal conception of moral judgment might look like. They do, very broadly, all think that emotions are terribly important, but beyond this three noticeably different views are articulated. To the extent that my own views have represented the advocacy of an opposing more cognitivist position, I haven’t denied the importance of emotions but have maintained that cognitive components are vital too (cognitive components, that is, for which the anti-nativist proposals do not succeed in accounting). Yet if asked to characterize the crucial cognitive elements of the more demanding conception, I have nothing so simple and distinct as “recursion” to say. Rather, I will point to aspects of moral judgment like *desert*, *transgression*, *practical authority* (etc.), and declare (A) that these are *cognitions* (e.g., judging that X deserves punishment is not something one just “feels”), and (B) that emotional resources alone do not suffice to account for them. But the answer lacks precision (though is no less reasonable for that): The list of cognitions is worryingly open-ended (note the “(etc.)”) and, moreover, not one of the items listed is easily defined. The literature on with what kind of “practical authority” our moral norms are invested, for example, stretches back to the ancient Greeks and continues unabated.

Perhaps my pessimism is premature and distinct senses of “moral judgment” can be delineated with a reasonable amount of specificity. Or perhaps my doubt will be borne out, and the whole concept will remain inchoate and ill-defined. In either case, what is evident is that it is a mistake to choose one particular characterization of “moral judgment” and declare it to be the true and unique deserver of that name. I have argued elsewhere (Joyce 2012) that this kind of indeterminacy may span the difference between metaethical cognitivism and noncognitivism, and also the difference between moral realism and moral skepticism. In other words, there may be some legitimate precisification of the concept *moral rightness* (for example) according to which rightness is a real property of certain actions; but there may be other equally legitimate precisifications according to which no such property exists anywhere. How might this sort of indeterminacy affect the debate over moral nativism?

It is possible (and not unlikely) that on *any* precisification of “moral judgment” (and on any disambiguation of “innate”) moral nativism is false. But it is also possible that moral nativism is true for certain precisifications and false for others. Certainly the *plausibility* of various pro-nativist and anti-nativist arguments varies according to different conceptions of the target trait. For example, if one is concerned with questions of universality, then the less demanding is our conception of a moral judgment, the more likely it is that we will find evidence of universality, since, as a truistic rule of thumb, X+Y is going to occur more often than X+Y+Z. On these grounds Stich objects that the rich conception of moral judgment that I offered (in Joyce 2006) spells problems for moral nativism: “For if moral judgment requires *all of that*, what reason is there to think that people in cultures very different from ours *make*

moral judgments?” (2008, 233).<sup>15</sup> If this is correct, then (roughly) richly-construed moral judgments are less likely to be universal, thus favoring the non-nativist case (various aforementioned complications with universality aside).

A number of opponents of moral nativism allow that some kind of *normative* nativism might be true. Earlier we saw Nichols accept nativism about non-hypothetical norms. Edouard Machery and Ron Mallon (2010) also accept the plausibility of nativism about normative cognition (“that is, the capacity to grasp norms and to make normative judgments” (4))—where nativism is understood in evolutionary terms. What they insist upon is that *moral* judgment is but a proper subset of the normative, and there is no evidence for any psychological adaptations dedicated to moral thinking in particular. While it cannot be reasonably denied that the category of the normative is larger than the category of the moral, it should also be noted that *how much larger* depends on what conception of the moral one endorses. A demanding conception will make the moral a smaller subset of the normative; a less demanding conception will yield a larger subset. The larger the subset, however, the more plausibility there is to the claim that it is in fact *moral* judgment that is the distinct adaptation, while the human capacity to make non-moral normative judgments is a case of aspects of a biological adaptation being coopted for new uses.<sup>16</sup> (This position will be strengthened if we have a plausible hypothesis about why moral judgment in particular might have been adaptive to our ancestors while lacking a hypothesis about why normative judgments in general might have been adaptive.) Thus, again, a less demanding conception of moral judgment might be more amenable to a nativist explanation than a more demanding one.

On the other hand, POS arguments seem to cut the other way. If a thin moral judgment can be constructed out of evolutionarily pre-existing mechanisms, then heaping more demands on the conception of moral judgment (“thickening” it) lowers the probability that these mechanisms will remain sufficient to the explanatory task. Again speaking roughly: Richly-construed moral judgments will need more mechanisms to explain them; and the more mechanisms to which one must appeal, the more likely it is that at some point one will need to appeal to a *dedicated* mechanism, thus favoring the nativist case. In this chapter I haven’t attempted the difficult task of arguing that a POS-style argument is plausible even for a demanding conception of moral judgment (though I admit to some sympathy with the project); my objective is simply to draw attention to the fact that the plausibility of the argument may vary according to how the target trait is drawn.

## Conclusion

The upshot is that both moral nativism and moral non-nativism may be perfectly defensible positions, and may remain so even when all data are in. This, I predict, will not be a popular

---

<sup>15</sup> Machery and Mallon (2010) make the same point: “Joyce’s claim is substantive and provocative precisely because of the rich characterization of moral judgments that he offers.”

<sup>16</sup> I am making a debatable background assumption here: that if trait *T* has adaptive function *F<sub>a</sub>*, then, for whatever processes make possible “coopting” *T* for new functions *F<sub>b</sub>*, *F<sub>c</sub>*, etc., it will be *prima facie* more probable that these processes will have coopted *T* for fewer new functions than for more new functions. Assessing such a principle would be a complicated task; here I leave it at an intuitive level.

conclusion—philosophers and scientists alike prefer their truths tidier—but it is surely worthwhile to diagnose, in advance, those points of conceptual imprecision that may confound future debate.<sup>17</sup>

#### REFERENCES:

- Ariew, A. 1996. "Innateness and canalization." *Philosophy of Science* 63 (suppl.): S19-S27.
- Ariew, A. 1999. "Innateness is canalization: In defense of a developmental account of innateness." In *Where Biology Meets Psychology: Philosophical Essays*, ed. V. G. Hardcastle, 117-138. Cambridge, MA.: MIT Press.
- Blackburn, S. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Blackburn, S. 1998. *Ruling Passions*. Oxford: Oxford University Press.
- Buller, D. 2006. *Adapting Minds*. 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 *Mind and Cognition*, ed. W. Lycan, 627-646. Oxford: Blackwell.
- Cosmides, L. & Tooby, J. 1992. "Cognitive adaptations for social exchange." In *The Adapted Mind*, ed. J. H. Barklow, L. Cosmides, & J. Tooby, 163-228. Oxford: Oxford University Press.
- Cowie, F. 1999. *What's Within: Nativism Reconsidered*. Oxford: Oxford University Press.
- Darwin, C. 1879/2004. *The Descent of Man*. London: Penguin Books.
- Darwin, C. 1877. "A biographical sketch of an infant." *Mind* 2: 285-294.
- Dwyer, S. 2009. "Moral dumbfounding and the linguistic analogy: Methodological implications for the study of moral judgment." *Mind & Language* 24: 274-296.
- Dwyer, S., Huebner, B., & Hauser, M. 2010. "The linguistic analogy: Motivations, results, speculations." *Topics in Cognitive Science* 2: 486-510.
- Foot, P. 1972. "Morality as a system of hypothetical imperatives." *Philosophical Review* 81: 305-316.
- Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. 2001. "An fMRI investigation of emotional engagement in moral judgment." *Science* 293: 2105-2108.
- Griffiths, P. 2002. "What is innateness?" *Monist* 85: 70-85.
- Haidt, J. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment." *Psychological Review* 108: 814-834.
- Haidt, J. & Joseph, C. 2004. "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues." *Daedalus* 133: 55-66.
- Hamilton, W. 1964. "The genetical evolution of social behavior" I and II. *Journal of Theoretical Biology* 7: 1-52.
- Hauser, M. 2006. *Moral Minds*. NY: Harper Collins.

---

<sup>17</sup> Thanks to my co-editors for valuable feedback. Some of the passages in this paper concerned with psychological altruism are lifted more-or-less verbatim from my entry on "altruism and biology" for the *International Encyclopedia of Ethics*.

- Hauser, M., Chomsky, N., & Fitch, W. T. 2002. "The faculty of language: What is it, who has it, and how did it evolve?" *Science* 298: 1569-1579.
- Heine, B. 1985. "The mountain people: Some notes on the Ik of north-eastern Uganda." *Africa: Journal of the International African Institute* 55: 3-16.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2006a. *The Evolution of Morality*. Cambridge, MA.: MIT Press.
- Joyce, R. 2006b. "Is human morality innate?" In *The Innate Mind: Culture and Cognition*, ed. P. Carruthers, S. Laurence, & S. Stich, 257-279. New York: Oxford University Press.
- Joyce, R. 2009. "Review of Jesse Prinz's *The Emotional Construction of Morals*." *Mind* 118: 508-518.
- Joyce, R. 2011. "The error in 'The error in the error theory.'" *Australasian Journal of Philosophy* 89: 519-534.
- Joyce, R. 2012. "Metaethical pluralism: How both moral naturalism and moral skepticism may be permissible positions." In *Ethical Naturalism: Current Debates*, ed. S. Nuccetelli & G. Seay, #. Cambridge: Cambridge University Press.
- Knight, J. 1994. "The Mountain People as tribal mirror." *Anthropology Today* 10: 1-3.
- Machery, E. & Mallon, R. 2010. "The evolution of morality." In *The Moral Psychology Handbook*, ed. J. Doris, G. Harman, S. Nichols, et al., 3-46. Oxford: Oxford University Press.
- Mackie, J. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin Books.
- Mameli, M. 2008. "On innateness: The clutter hypothesis and the cluster hypothesis." *Journal of Philosophy* 105: 719-737.
- Mameli, M. & Bateson, P. 2007. "The innate and the acquired: Useful clusters or a residual distinction from folk biology?" *Developmental Psychobiology* 49: 818-831.
- Mikhail, J. 2008. "The poverty of the moral stimulus." In *Moral Psychology Volume 1: The Evolution of Morality: Adaptations and Innateness*, ed. W. Sinnott-Armstrong, 353-360. Cambridge, MA.: MIT Press.
- Mill, J. S. 1861. "Utilitarianism" chapter 3. In *Fraser's Magazine*, November: 525-534.
- Nichols, S. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.
- Nichols, S. 2005. "Innateness and moral psychology." In *The Innate Mind: Structure and Contents*, ed. P. Carruthers, S. Laurence, & S. Stich, 353-430. New York: Oxford University Press.
- Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Prinz, J. 2008a. "Is morality innate?" In *Moral Psychology Volume 1: The Evolution of Morality: Adaptations and Innateness*, ed. W. Sinnott-Armstrong, 367-406. Cambridge, MA.: MIT Press.
- Prinz, J. 2008b. "Resisting the linguistic analogy: A commentary on Hauser, Young, and Cushman." In *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, ed. W. Sinnott-Armstrong, 157-170. Cambridge, MA.: MIT Press.
- Prinz, J. 2009. "Against moral nativism." In *Stich and his Critics*, ed. D. Murphy & M. Bishop, 167-189. Malden, MA.: Wiley-Blackwell.

- Prinz, J. Forthcoming. "Where do morals come from? – A plea for a cultural approach." In *Empirically Informed Ethics*, ed. M. Christen, M. Huppenbauer, C. Tanner, & C. van Schaik. Springer Press.
- Schino, G. & Aureli, F. 2010. "The relative roles of kinship and reciprocity in explaining primate altruism." *Ecology Letters* 13: 45–50.
- Sober, E. 2000. "Psychological egoism." In *The Blackwell Guide to Ethical Theory*, ed. H. LaFollette, 129-148. Oxford: Blackwell.
- Sober, E. & Wilson, D. S. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
- Sripada, C. S. 2008. "Nativism and moral psychology: Three models of the innate structure that shapes the contents of moral norms." In *Moral Psychology Volume 1: The Evolution of Morality: Adaptations and Innateness*, ed. W. Sinnott-Armstrong, 319-343. Cambridge, MA.: MIT Press.
- Sripada, C. & Stich, S. 2006. "A framework for the psychology of norms." In *The Innate Mind: Culture and Cognition*, ed. P. Carruthers, S. Laurence, & S. Stich, 280-301. New York: Oxford University Press.
- 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.
- Stich, S. 2008. "Some questions about *The Evolution of Morality*." *Philosophy and Phenomenological Research* 77: 228-236.
- Tangney, J. & Fischer, K. (ed.). 1995. *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment, and Pride*. New York: Guilford Press.
- Turnbull, C. 1972. *The Mountain People*. New York: Simon & Schuster.
- West, S., Griffin, A., & Gardner, A. 2007. "Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection." *European Society for Evolutionary Biology* 20: 415-432.