Philos Stud (2009) 144:189–209
DOI 10.1007/s11098-008-9205-5

Moral judgment purposivism: saving internalism from amoralism

M. S. Bedke

Published online: 9 February 2008

Abstract Consider orthodox motivational judgment internalism: necessarily, A's sincere moral judgment that he or she ought to φ motivates A to φ. Such principles fail because they cannot accommodate the amoralist, or one who renders moral judgments without any corresponding motivation. The orthodox alternative, externalism, posits only contingent relations between moral judgment and motivation. In response I first revive conceptual internalism by offering some modifications on the amoralist case to show that certain community-wide motivational failures are not conceptually possible. Second, I introduce a theory of moral motivation that supplements the intuitive responses to different amoralist cases. According to moral judgment purposivism (MJP), in rough approximation, a purpose of moral judgments is to motivate corresponding behaviors such that a mental state without this purpose is not a moral judgment. MJP is consistent with conceptual desiderata, provides an illuminating analysis of amoralist cases, and offers a step forward in the internalist-externalist debates.

Keywords Internalism · Externalism · Evolutionary ethics · Amoralism · Cognitivism · Expressivism · Noncognitivism · Proper function

If John Doe sincerely judges that he ought to visit his grandmother in the hospital, we expect him to be so motivated, at least a little bit. If Jane Doe sincerely judges that she ought to help someone in dire need (at no inconvenience to herself), we expect her to be so motivated, again, at least a little bit. And our expectations in

M. S. Bedke (✉)
Department of Philosophy, University of Arizona, Social Science Bldg., Rm 213,
Tucson, AZ 85721, USA
e-mail: mbedke@email.arizona.edu

these and similar cases are usually met, for first person moral judgments[1] are usually motivationally efficacious, though the motivation can be outweighed by other considerations. That much seems uncontroversial, but there are different ways of clarifying and articulating the connection between moral judgment and motivation. The orthodox[2] internalist position, motivational judgment internalism, is that there is an a priori conceptual necessity between moral judgments and motivation along the following lines:

> MJI: necessarily, individual A's moral judgment that he or she ought to φ provides A with at least some motivation to φ.

Given that A has a mental state M that bears all other markings of a moral judgment but lacks motivational oomph, MJI holds that M *cannot* be a genuine moral judgment. This view strains the thought that amoralists—individuals who acknowledge moral obligations but remain unmoved by them—are at least conceptually possible. The orthodox externalist alternative says there is no a priori conceptual connection between moral judgments and motivation. Any observed motivational regularities are merely contingent. Insofar as the externalist line accommodates the possibility of amoralists it would seem to have the upper hand.

In response to this debate I want to, first, open some middle ground between these extremes by reflecting on different kinds of amoralist cases. I present cases of community-wide amoralism that underwrite a conceptually grounded connection between moral judgments and motivation while respecting the conceptual possibility of *isolated* amoralists, or amoralists embedded in communities of widespread moral motivation. Secondly, I consider whether empirical information, particularly information about the evolutionary history of our moral practices, might underwrite a more robust connection between moral judgments and motivation than we find in orthodox externalism, and whether this empirical information might underwrite a synthetic necessity principle narrower than the a priori connection found by reflecting on cases. To these ends, I introduce and defend moral judgment purposivism (MJP), which can be provisionally stated as follows:

> MJP: the *purpose* of a moral judgment is to motivate individuals to act accordingly.

As we shall see, MJP respects most intuitions regarding the possibility of various kinds of amoralists, provides a more determinate conception of the connection between moral judgments and motivation, and best articulates morality's essential action-guiding role. In closing we shall see what all this means for the traditional

[1] Throughout I will be concerned with first person moral ought judgments. The literature on this kind of internalism is large. See Darwall (1983), Schaffer-Landau (2003), Ch. 6, and Mele (1996, 2003) for nice discussions.

[2] Some with internalists leanings could retreat to weaker and less orthodox claims, like 'Necessarily, the virtuous are motivated by their moral judgments' or 'Anyone who judges that he ought to X (morally or otherwise) is either motivated accordingly or practically irrational.' But such theses are more about virtuous or rational people than they are about moral judgments. These theses require the motivational connection to be mediated by something external to the judgments themselves. I want to argue that there is a less mediated motivational role for moral judgments that denies the robustly externalist view that moral judgments *of themselves* have only contingent connections to motivation.

arguments from internalism to non-cognitivism, and from externalism to non-cognitivism.

1 The isolated amoralist

An amoralist is one who makes moral judgments about his or her moral obligations, but utterly fails to be moved by them. Brink (1989, pp. 45–50) forcefully argues for the conceptual possibility of amoralists,[3] and recent empirical work supports Brink's position. In a study by Shaun Nichols subjects are given a description of a psychopath who acknowledges that hurting others is morally wrong, but who claims not to care, and who in fact hurts others. Nichols discusses the results as follows:

> Most subjects maintained that the psychopath did really understand that hurting others is morally wrong, despite the absence of motivation .... *Prima facie*, this counts as evidence against the [internalist's] inverted-commas gambit. For it seems to be a platitude that psychopaths really make moral judgments. And if it is a platitude that psychopaths really make moral judgments, it will be difficult to prove that [internalism] captures the folk platitudes surrounding moral judgment (Nichols 2004, pp. 74–75).

Widely shared intuitions in these cases suggest that, at least insofar as the concept of a moral judgment is concerned, amoralists are possible.

These amoralist cases are troubling for orthodox internalism and they are taken to support orthodox externalism in that they seem to show that motivation is no part of the concept of a moral judgment.[4] However, there are ways to press back against orthodox externalism. Consider the views of Mark Timmons, James Dreier, and Simon Blackburn, who suggest a connection between moral judgments and motivation looser than MJI, but stronger than externalism. Timmons (1999), for instance, claims that "a moral judgment typically has certain (defeasible) causal tendencies, including, especially, certain first-person choice guiding tendencies .... *Its typically having these tendencies is part of the very concept of moral judgment*" (Timmons 1999, p. 140, emphasis added). Similarly, Dreier (1990) says, "let us call modest internalism the principle that in normal contexts a person has some motivation to promote what he believes to be good" (Dreier 1990, p. 14). And Blackburn remarks:

[3] Not only does MJI rule out amoralists, it also rules out the possibility that a normally virtuous person on one occasion renders a moral judgment that fails to motivate. The natural thing to say is that these are motivationally *defective* moral judgments, though still moral judgments. Proponents of MJI try to avoid the natural thing to say. They might say that the amoralist only makes a moral judgment in the *inverted commas* sense; that the moral judgment is not *sincere*; or that the moral judgment does not concern a moral *obligation*. These moves might suffice for the amoralist, but they seem quite desperate explanations of the single-shot glitch in the otherwise virtuous person.

[4] Nichols (2004, p. 111) embraces what he calls "empirical internalism about core moral judgment". In Nichols's words, "core moral judgment is nomologically connected with motivation" (Id). Though he uses the label "internalism," this appears to be the standard externalist view advocated by Brink, for any connection between moral judgment and motivation would be contingent.

> My own judgment on this debate is that externalists can win individual battles. They can certainly point to possible psychologies about which the right thing to say is that the agent knows what it is good or right to do, and then deliberately and knowingly does something else. And they can point to psychologies like that of Satan, in which it can become a reason for doing something precisely that it is known to be evil. But internalists win the war for all that, in the sense that these cases are necessarily parasitic, and what they are parasitic upon is a background connection between ethics and motivation. They are cases in which things are out of joint, but the fact of a joint being out presupposes a normal or typical state in which it is not out. (Blackburn 1998, p. 61).

There are at least two ways one could understand the typical, normal, or background connection alleged in these passages. One could read it either as a property of communities, or as a property of individual psychologies. According to the latter view, if we know that a particular agent, A's, tokenings of mental state M do not typically or normally have the tendency to motivate A, then we know the tokens of M are not genuine moral judgments. This is the less plausible view, for it rules out lifelong amoralists, which seem conceptually possible. That is, an agent A might render genuine moral judgments that never have motivational force for A.[5] The better view is that something like typicality or normalcy or the background connection is a property of *communities* of agents such that amoralists, even lifelong ones, render moral judgments so long as their community members typically or normally render motivationally efficacious moral judgments. To adequately support this view I want to reflect on various amoralist cases that vary in the kind and degree of community-wide motivational tendencies.

2 Cases of community-wide amoralism

Test cases will describe various kinds of community-wide connections between moral judgment and motivation. Reflection of these cases will help us determine whether isolated amoralist should indeed be treated differently from communities of amoralists, and whether we can thereby ground some kind of pre-theoretic, analytic connection between moral judgments and systematic motivational tendencies. Let me begin by addressing the shortcomings of some other test cases before introducing the probative case of Amoralsville.

[5] Both Dreier and Blackburn seem to thinking of normalcy or the background connection as a community property, though this is not entirely clear. For example, it is implicit in Dreier's discussion of Sadists, who individually do not have the slightest motivation to perform moral actions but nonetheless render moral judgments, that normalcy is a property of a population. On the other hand, he elsewhere suggests that psychic states can defeat normalcy, (e.g., he says that "if a person has a certain state of character for all her life, then behavior flowing from that state is normal for her" (Dreier 1990, p. 14)), which indicates that normalcy can be a property of particular psychologies.

2.1 Amorality and Alpha

James Lenman (1999) offers a starting point for thinking about widespread amoralism. He asks us to consider an entire planet of amoralists, Amorality, where we are told scientists ascertain and record moral facts, but where no one is ever practically motivated by their moral judgments (1999, pp. 445–446). Lenman finds this hypothetical society “preposterous,” and uses it to support the view that global amoralism is impossible. He adds to this the premise that if global amoralism is not possible, then neither are isolated amoralists amongst us and with our background of motivational efficacy. Hence externalism is false.

Unfortunately, Lenman does not provide much more detail about the nature of this society on Amorality. We are told that the residents study moral facts—indeed that there are morality departments and moral scientists—but we are not told much about what it is like living on Amorality, and we are left wondering how a society could hold itself together with absolutely no moral motivation. Assuming some of these details can be provided to make the hypothetical more cogent, there are problems both with Lenman’s reaction to Amorality, and his inference from the premise that entire communities of amoralists are not possible to the conclusion that externalism is false. First, what seems to bother Lenman about the putative moral judgments on planet Amorality is their lack of any *propositional content* (1999, pp. 452–453). As he understands it, the propositional content of putative amoralist judgments are parasitic on the judgments of moralists, for putative amoralist judgments are really made in the inverted commas sense; they are judgments about the judgments moralists make. On Amorality there are no moralists and there never were, so the putative amoralist judgments lack content, or so the argument goes.

This argument can be challenged in various ways. But assuming *arguendo* that it is sound, it fails to capture the point of disagreement between internalists and externalists. That disagreement is not supposed to be about content and whether amoralist moral judgments are propositionally meaningless, but rather whether moral judgments with agreed upon contents must have motivational import. To better test externalism as applied at the level of global motivational failure, we would do better to think of a case where the semantic contents of the amoralist judgments are not at issue. In addition, even if Lenman’s scenario is impossible, the move from the impossibility of global amoralism to the impossibility of isolated amoralism is under-motivated. Perhaps some background motivation is necessary to either breath semantic content into the judgments of amoralists, or to otherwise understand the judgments they make as moral judgments, but it would not follow that isolated amoralists in our community fail to render genuine moral judgments.

To see whether community-wide amoralism is conceptually impossible *on motivational grounds* we should construct a case that is similar to our moral situation, save some community-wide lack of motivation. Noting the above difficulties with Lenman’s case, Gert and Mele (2005) try to further guide the way. They ask us to consider planet Alpha, where beings “emerge with a strong genetic predisposition to acquire generic desires to do whatever they morally ought” (277). They imagine that a worldwide catastrophe strikes that sinks all the residents of Alpha into a deep listlessness. Though the residents continue to make genuine first person moral ought judgments, these judgments no longer carry motivational oomph, for everyone is caught in the grip of a depressive funk. Gert and Mele think that this kind of scenario is possible, and I suspect many will share that intuition. If so, this case calls into question our working hypothesis—that some community-wide typical, normal, or otherwise background moral motivation is needed for individuals to render moral judgments.

Certainly this case is evidence that a *certain kind* of community-wide amoralism is possible, and from there it is tempting to conclude that there need not be any community-wide background connection between moral judgment and motivation, or some sense of normalcy or typicality in moral motivation. But that would be too quick. What is missing in Alpha is a *present statistical* background connection between moral judgment and motivation. Perhaps that is not essential to moral judgments, but there is another sense in which Gert and Mele's case exhibits a background connection. As the evolutionary tale emphasizes, the moral judgments on Alpha have conduct guidance as part of their *evolutionary function*, so we might understand the background connection in terms of evolutionary, historical considerations. If the listlessness persists, one wonders whether we would be less and less inclined to consider those judgments genuinely moral as the historical function of moral judgments fades from view.

A better test case would eliminate this historical, functional background connection between moral judgments and motivation. I now propose a case that does just that, and that supports the view that a certain kind of community-wide amoralism is conceptually impossible.

2.2 The amoralsville case

Consider a distant community very much like our own, but unrelated to our own (perhaps they are on another planet) that developed a very stringent, heavy-handed system of punishment and coercion to keep its citizens in line. The residents of this community are ruled by a single dictator that metes out severe punishments, but only for behaviors that by and large happen to violate our ethical norms. As a result, individuals in this community generally keep their contracts, respect each other's property, and help those in need because they fear punishment and coercion should they fail to do so. As external observers we would say that their behaviors by and large conform to our ethical norms, though we realize that they are never motivated by anything other than their own interests and fear of harm to their interests.

Let us call this place *Amoralsville*. So far, there are not even putative moral judgments in Amoralsville. But imagine that the residents receive radio frequencies from our community and thereby observe our use of moral language and discourse. With the introduced moral vocabulary, the residents of Amoralsville learn to apply moral concepts correctly. As a result, Amoralsville residents correctly pick out what is right and wrong, acknowledge obligations, and can correctly categorize that which they (morally) ought to do. In fact, forming first personal putative moral judgments and speaking in ethical terms becomes kind of a fad in Amoralsville, though the judgments never garner any motivational force, and moral demands simply do not weigh with them. Residents of Amoralsville are at all times solely motivated by their own interests.

Because this community learns whatever propositional meaning there is to moral language, and uses it correctly to classify cases, there is no objection that their judgments lack content, which is an improvement over Lenman's Amorality case. Though we are not assuming that the Amoralsville residents learn the psychological and social roles that moral language plays in our earthly communities, such highly theoretical knowledge is not needed for linguistic competence (note that most of us do not know such roles, hence the internalism debates, yet presumably most of us are competent users of the language). The community is in most respects relevantly like our own, including the kinds of behaviors typically engaged in, except moral judgments do not now, nor did they ever, perform any kind of social function, which serves to distinguish Gert and Mele's Alpha case. The question is: Do Amoralsville residents render genuine moral judgments? This is quite a different case than those involving isolated amoralists. It looks like the citizens of Amoralsville do not really engage in genuine ethical discourse. Intuitively, an essential ingredient of ethical discourse has gone missing, viz., its action-guiding character. When actual and historical motivational functions go missing in a community its members do not render moral judgments. Thus, some background connection between moral judgments and motivation is necessary.

This is intuitive evidence that the orthodox externalist line is mistaken, but we do not thereby vindicate orthodox internalism. The Amoralsville case supports some middle ground internalist thesis, but I fear that intuitions on cases like this can only get us so far. It is not clear that more fine grained Amoralsville-type cases concerning degrees of community moral motivation and kinds of background function will deliver clear intuitions with probative value. In any event, rather that chase that lead, I now consider whether empirical information can supplement the case for a middle ground internalist thesis. Below I defend a three part hypothesis that capitalizes on the relevance of historical considerations: (1) Our concept of a moral judgment picks out a type of mental state that is embedded in wider social practices with evolutionary histories, (2) this evolutionary history provides moral judgments with a purpose to motivate certain behaviors, and (3) this purpose is (or should become) a necessary feature of moral judgments. This hypothesis will respect the a priori intuitions on cases given above and provide a more determinate conception of the link between moral judgments and motivation.

3 A purpose of moral judgments

I argue that moral judgments are best understood as part of larger moral practices within communities of moral agents, and that situating moral judgments in these larger practices best illuminates their motivational character. What follows will be somewhat exploratory and suggestive. I rely on some recent thoughts in evolutionary theory, and I grant that my suggestions will be subject to further study, which I take to be a virtue. In any event the discussion opens up new possibilities to advance the traditional internalist–externalist debate and makes a good case that there are interesting connections between morality and motivation that are not contingent (as traditionally understood by the externalist).

Section 3.1 articulates a biological theory of purposes, or proper functions. According to the view, natural objects can acquire purposes by virtue of the selection processes that occur during evolution. Section 3.2 then applies this view about biological purposes to moral practices to show that moral judgments also have evolved purposes, one of which is to motivate prototypically moral behaviors. After laying down the principles of MJP in Sect. 3.3 and anticipating some objections, Section 4 shows that the view respects our considered judgments about isolated amoralists and the residents of Amoralsville. More than this, the purposive view offers us a way of looking at these cases that captures and clarifies the differences between them, provides a conception of what it is for moral judgments to normally or typically motivate (or, as I shall prefer to say it, moral judgments are *supposed to* motivate, or it is their *purpose* to motivate), and thereby further informs our concept of a moral judgment.

3.1 Evolved purposes

The purpose of our moral judgments is just one instance of a general theory of the purposes of evolved functionings. Consider the familiar case of genetic replication. Various genes express themselves as phenotypes, and the phenotypic expression of a given kind of gene can make that gene a more or less successful replicator depending on how the phenotype functions within an environment. Given an environment where various genes express various phenotypes, those phenotypes that increase a gene's relative rate of replication will count as adaptive and so increase the proportion of the gene generation after generation.

From this basic story Ruth Millikan (1984) has developed a theory of *proper functions* that applies to biological entities and (she argues) languages. Her story will closely parallel our account of the purposes of moral practices. She considers things like the human heart and asks, what makes a heart the kind of thing that it is? Millikan gives the following partial reply: hearts are things that have pumping blood as a proper function. Very roughly we can say that a function F is a proper function of an entity if F made that entity's ancestors selectively fit, and so caused the entity's ancestors to proliferate relative to its competitors. We can explain why some entities exist today by appealing to the way in which those kinds of things functioned historically, and the functions we appeal to in these explanations are proper functions. It turns out the pumping blood was and still is a useful function for biological organisms to have, and so hearts were selected for, and the genes that expressed them were more likely to replicate, precisely because they performed that function.[6]

[6] One interesting question is whether the heart's proper function is partially definitive of the kind of thing that it is. There is a good case that it is. Surgically removed hearts, defective hearts, and other things that do not pump blood might still count as hearts depending on their proper functionings, which, in turn, depend on the histories of these things' ancestors. But water pumps do not count as hearts even if they can pump blood precisely because, it would seem, water pumps do not have pumping blood as a proper function, and having such a proper function is partly definitive of hearts. This issue will come to the fore below in a discussion of MJP as a synthetic necessity claim.

It is natural to use ‘purpose’ to capture what we mean by a selected proper function, as we might say that the purpose of a heart is to pump blood, and so we can shorthand this complicated story of selection and propagation through time by referring to an entity’s purposes.

3.2 Moral evolution

While the adaptationist story is familiar in biology, it has a very general form. We can model change through time with evolutionary dynamics if: (1) there is a population with varied phenotypes, where (2) the phenotypes are copiable, and (3) different phenotypes result in differential relative copying success. More importantly, if these three conditions obtain for any phenotype we *expect* some evolutionary model to explain why the phenotype was selected. And if some phenotype was selected for, then it has a corresponding proper function or purpose.

I want to suggest that moral judgments inherit a purpose by playing a role in certain wider moral practices, so let me begin by considering whether prototypical cases of moral behaviors fit these three conditions for adaptationist modeling. Different behaviors have functional differences that can be copied by others or copied on other occasions and the behaviors can impact one’s relative fitness. Consider a classic case—the prisoner’s dilemma—and suppose the prisoners are trying to determine whether to keep or break their prior promise to cooperate. In this case, defectors can take advantage of those who keep their promise, and dominance reasoning actually recommends that each party defect no matter what the other party does. This case presents an adaptationist puzzle that is reiterated for many moral behaviors, viz., why would a habit of keeping one’s promises evolve, given its continual vulnerability to the defection strategy? Can we explain why behaviors such as refraining from theft, lying, and intentionally harming others would evolve given the seemingly obvious fitness payoffs for contrary behaviors? Can we explain why honesty and helping behaviors would evolve?

The answer to these questions is ‘yes.’ If fact, the problem now is not whether these behaviors can be explained, but which explanation is the best one. Hamilton (1963) introduced the idea of inclusive fitness, or kin selection, which can explain why genetically related individuals might help each other out. Trivers (1971) expanded the idea by discussing reciprocal altruism, where individuals who help only those others who reciprocate can gain a fitness advantaged over non-reciprocators, who are left to fend for themselves. In both cases, genetically altruistic behavior given to a non-kin non-reciprocator puts one at a fitness disadvantage, but so long as the altruistic behavior is correlated with other altruistic behavior to a sufficient degree, genetic altruists are expected to proliferate.[7] Sober and Wilson (1998) categorize these kinds of theories as models of group selection and they survey other scenarios wherein altruistic behavior can evolve. Brian Skyrms (1996, 2004) has produced models showing the evolutionary stability of

[7] For a review of game theory models of social behaviors see Maynard Smith (1982) and Axelrod and Hamilton (1981).

mutual aid, respecting property, some forms of punishment behavior, and other cooperative behaviors. As with the foundational work of Hamilton and Trivers, in these models a factor that permits prototypically moral behaviors to evolve is *assortive interactions*, or the ability of moral actors to interact with one another (rather than non-moral actors) a sufficiently high proportion of the time. Of course, the view here is consistent with the presence of other, non-assortive mechanisms, though historically the evolution of morality has been most puzzling for failure to appreciate these models. Moral actors as a class can become fitter than non-moral actors when they stick to their own kind and avoid too much free riding and predation.

Though the basic story is simple, it is important to be sensitive to some nuances. We should, for example, make the conceptual distinction between the unit that replicates (the replicator) and the unit upon which selective pressures proximately act (the interactor).[8] To see how these can come apart imagine that some genetic material replicates from one generation to the next, and in each generation the genetic material fully determines some moral behavior or behaviors that are selected for in favorable social environments. Here it is clear that the unit of replication, genetic material, is distinct from the unit of interaction, moral behavior. Focusing on replicators for the moment, another possibility is that moral behaviors themselves are non-genetically replicated via cultural processes.[9] Perhaps most plausible are hybrid views where genes and culture play a replicator role. One such view has been defended by psychologists Haidt and Joseph (2004, 2007), who argue that relevant studies on human and primate moral behavior evidence five innate moral modules concerning suffering/compassion, reciprocity/fairness, hierarchy/respect, purity/disgust, and ingroup/outgroup loyalty attitudes.[10] Different cultures can then emphasize different aspects of these moral sentiments, but the innate mechanisms will limit the kinds of moral practices people can adopt. This hypothesis has the virtue that it can explain the common structure and some of the common intuitions shared by moral practices and at the same time explain the variations we actually find across cultures (cf. Haidt 2001). So long as we remain sensitive to these details, and to the real possibility that genetic replication and cultural replication can pull in different directions,[11] we can gain insight by idealizing a bit to model the processes at work.

Replicators and interactors are also conceptually distinct from units that *benefit* from adaptation. In many of the above models moral individuals need not benefit, either in the sense of being a successful reproducer or in a more intuitive sense, for often moral behavior requires individual sacrifice. In some sense populations or species seem to benefit by having individuals that exhibit moral behaviors, for groups with many moral actors do better (in terms of persisting through time) than groups lacking such members. The models we have been considering explain all of

[8] These terms are due to Hull (1980).

[9] For a nice discussion of cultural evolution see Richarson and Boyd (2005).

[10] See also Haidt and Bjorkland (2008).

[11] For a nice discussion of some difficulties when modeling culturally driven adaptations see Dennett (1995, Ch. 12).

this by assuming that moral behaviors are the interactor units, though we should be open to the possibility of some genuine group selection, where groups or features of groups comprise interactor units that are not fruitfully reduced to some more basic interactor units.

Bearing in mind these nuances, the main claim here is that some such evolutionary story will provide the best explanation of (at least prototypical instances of) our moral behaviors precisely because moral behaviors fit the three conditions for applying evolutionary dynamics.[12] If this is right, then the historical ancestors of some of our current moral practices would have performed some function, and they were selected for, copied, and propagated precisely because they performed that function.

Applying Millikan's proper functionalism, our moral behaviors have a proper function, or a purpose, that corresponds to the functions for which they were selected. The prevailing evidence from biology and psychology indicates that (at least prototypical) moral behaviors evolved through a process of assortive interactions, which we might call social interactions because they enable cooperative, mutually advantageous outcomes amongst moral actors. Given the evidence we can claim as a working hypothesis that a purpose (or proper function) of moral behaviors is to enable and further cooperative, mutually beneficial outcomes for moral individuals.[13]

3.3 Moral judgment purposivism

So far we have talked about the evolution of moral behavior, though it would be more appropriate to talk about the evolution of moral practices, which includes behaviors, psychologies, language, and all else that helps individuals obtain mutually beneficial outcomes. In particular, to elicit certain moral behaviors we would have to have a mental state or states responsible for motivating the desired behaviors. Here it is helpful to think in terms of nested proper functions. Consider again the biological domain. One of a left ventricle's proper functions is to squeeze blood out of the heart and that is its contribution to the heart's overall function of pumping blood throughout the body. Similarly, individual psychologies will need mental states that play a role in translating recognitions of moral actions into behavior, thereby contributing their part to the moral practices that enable social cooperation. I propose that first person moral judgments were selected in part to play this role. Moral judgments with some connection to motivational states would

[12] Evolutionary dynamics seeks to explain the distal causes of things that copy over generations. Other approaches might provide workable proximate explanations of our moral behaviors so we can understand why individuals engaged in them from, say, a psychological or sociological perspective. But we need something like evolutionary dynamics to explain the persistence or proliferation of moral behaviors over generations.

[13] Our moral behaviors might fail to have these effects currently, but it is sufficient that their history bestows them with the purpose of eliciting these effects. More importantly, the mechanisms that enable our moral behaviors might have been co-opted to produce other kinds of behavior that fail to generate cooperative outcomes. I discuss this more below.

be selected for over moral judgments that merely recognize moral situations and obligations without helping to translate those into appropriate motivational states. Consequently, they would have a corresponding nested proper function: a purpose (or proper function) of moral judgments is to motivate individuals to act in accordance with the judgment. Note that purposes correspond to functions selected for, so we do not say that a purpose of moral judgments is to be a part of a wider social practice (that makes it unclear how it functions and why it was selected), but rather that a purpose of moral judgments is one of motivating behavior. This is not to deny the importance of other selected functionings in our moral practices, including the point and purpose of other-directed moral judgments in enforcing moral behaviors, but merely to point out a role (not *the* role) of first person moral judgments.[14]

There are various ways that moral judgments might play a motivational role. One possibility is that moral motivation, when it occurs, comes from moral judgments directly. Another possibility is that all motivational work is delegated to conative attitudes not considered part of the moral judgment per se.[15] On this second view one would have something like a *de dicto* conative attitude toward doing that which one morally ought to do, which would team up with cognitive moral judgments about what one ought to do to produce motivation. Though this issue cannot be fully addressed here,[16] I think prevailing evidence supports the first view.[17] Suffice it to say that even if moral judgments turn out to be motivationally inert in and of themselves, the core idea of the purposive view—that they have some role to play in generating motivation (if only to orient agents in the right direction)—remains intact.

We are now in position to set forth MJP (the necessity claim to be discussed shortly).

[14] Other evolutionary theorists draw a distinction, sometimes implicitly, between other-directed moral judgments, which can play some role in enforcing social norms, and self-directed moral judgments, which can play some role in ensuring that the judger complies with social norms. See, e.g., Gibbard (1990, Ch. 4); Joyce (2006, Ch. 4); Kitcher (2006). The present view speaks to self-directed moral judgments, and their psychological and social roles. I mean to acknowledge and leave open the enforcement role of other-directed judgments.

[15] When pressing me to clarify these options, an anonymous reviewer gives the following analogy. Just as a steering wheel orients a car in a certain direction, and the accelerator provides the spring of action, we can think of my moral judgment that I ought to φ as orienting me toward φing, with a separate conative attitude toward φing providing the spring of action.

[16] I defend the besire view in Bedke (ms).

[17] The separate motivational attitude view posits a *de dicto* desire to do the right thing (whatever that turns out to be), which looks like a moral fetish. As Michael Smith (1994, pp. 71–76) has noted, this depicts agents as caring about valuable things derivatively via a desire to do what one ought. From our own experience and what we know about moral others, this looks implausible. Even if I am wrong about this and we do aim at doing what is right and good, whatever that turns out to be, a plausible psychological picture must supplement these *de dicto* desires with *de re* desires, for we seem to be motivated by moral judgments about particulars. In addition, intuitions on the Amoralsville case show that certain community-wide motivational failures indicate a lack of moral judgments. We do not withdraw ascriptions of just any conative attitude, but moral judgments. Absent some explanation for why a failing external to moral judgments proper would induce us to withhold ascriptions of moral judgments, this is some evidence, albeit non-conclusive, that moral judgments include a motivational component.

> MJP: (necessarily) a purpose of an individual A's moral judgment that he or she ought to φ is to motivate A to φ, where there is a type of behavior T such that:
>
> a. A's moral judgment is part of a social practice whose purpose is, in part, to influence individuals to engage in T behavior, and
> b. T behavior is (prototypically) moral behavior.[18]

MJP is a (partial) theory about moral judgments that respects the conceptual guideposts given by intuitions on cases from Sect. 2. The present theory goes farther than the conceptual desiderata, yielding an empirically informed, more determinate conception of the background connection between actual moral judgments and motivation, one expressed in the language of purposes.

I also suggest that MJP be read as a synthetic necessity claim so that it is metaphysically impossible for a moral judgment to fail to have this purpose, much as it is metaphysically impossible for water to have a chemical composition other than H_2O. If the concept of a moral judgment really works like the concept of water, however, then with our evolutionary information in hand a-purposive moral judgments would seem as impossible as water composed of XYZ. To test this, consider a community that looks very much like our own except it just recently popped into existence (perhaps during a lightning storm in the swamps). The citizens of this community seem to talk and behave much like we do, but they are a-historical beings. In keeping with similar cases in the literature, let us call this a swamp community with swamp residents rendering swamp judgments. What of their putative moral judgments? They do not have the right history to satisfy MJP, and without being evolved or designed there is no sense in which their moral judgments have a *proper* function, or a purpose, as opposed to some merely *actual* function. (Compare: goose liver has the proper function of regulating goose metabolism, but it does not have pate ingredient as a proper function even though it actually functions as a pate ingredient). So if swamp citizens render genuine moral judgments, then motivational purpose is not necessary to moral judgments, MJP has a counterexample, and *actual* motivational function (community-wide) is a sufficient background connection for moral judgments. Based on the example, we may have to consider revising MJP so that various kinds of similarity to actual moral judgments, not just purposive similarity, ensures that the term and concept 'moral judgment' refers to swamp moral judgments. One option is this:

[18] I thank an anonymous reviewer for comments that reminded me of the importance of distinguishing the action of φing, considered as the intensional content of A's judgment, from the action considered more objectively as falling under certain act types, one of which satisfies conditions a and b. I take it that this resolves some potential difficulties when A is thinking of φing in such a way that it does not satisfy conditions a and b even though the judgment seems to have a motivational purpose. To discuss an example from the reviewer, suppose I judge that I ought to call Fred (to warn him of approaching danger). For the principle to apply, it is too much to require a social practice with the purpose of influencing individuals *to call Fred*. And we do not want to complicate my judgment to include its grounds, viz., that I ought to help others. Instead it is better to say that MJP applies when there is an act type, where the judgment is part of a social practice whose purpose is, in part, to influence individuals to engage in that type of action (here, a kind of helping behavior), that type enters into our adaptationist, purposive, explanations, and that type is prototypically moral.

> MJP revised: (necessarily) those mental states that are sufficiently similar to actual moral judgments are moral judgments, where a purpose of A's actual moral judgment that he or she ought to φ is to motivate A to φ, and where there is a type of behavior T such that:
>
> a. A's moral judgment is part of a social practice whose purpose is, in part, to influence individuals to engage in T behavior, and
> b. T behavior is (prototypically) moral behavior.

I do not want to retreat to a revised version just yet. Despite some intuition that swamp people render genuine moral judgments, I have reservations in relying on those intuitions, and I believe the best overall theory of moral judgments might require us to give these intuitions less credence. Let me suggest that as the historical function of actual moral judgments is further investigated, and as those findings become more widely appreciated, our intuitions about swamp cases should shift. No doubt shifts in reference can occur. Gareth Evans (1973) gives us the example of 'Madagascar,' which originally referred to some portion of mainland Africa, but when the term and concept was transmitted to Marco Polo, he used it to refer to the off-coast island that is now the referent of the term. Putnam (1988) gives us a more radical example where the definition (and the putative referent) of 'momentum' changed with the shift from Newtonian mechanics to Einstein's theory of relativity. I am suggesting a far less radical shift whereby a concept can undergo some *augmentation* and *precisification* as a result of empirical information. If some such story applies to 'water,' we might imagine that prior to empirical discoveries about the internal chemical composition of watery stuff, 'water' referred to both H_2O and XYZ, and indeed all kinds of watery stuff. And as a result of empirical discoveries the concept changed (while remaining the same concept) to a narrower intension and extension so that it no longer referred to XYZ. This current proposal deviates from traditional causal regulation theories and intensional theories whereby 'water' *never* referred to XYZ because 'water' has only referred to the essential internal nature of the watery stuff of our acquaintance, which happens to be H_2O. Against these traditional accounts, imagine an alternative history where there was no common internal nature to all watery stuff. In that case would not 'water' refer to all watery stuff, including XYZ? What do we gain by saying that water nevertheless tries to refer to *the* essential internal nature of the watery stuff of our acquaintance, only to fail for lack of a uniform internal nature? In this vein consider milk, which does not enjoy a uniform internal nature, but does enjoy shared functional and etiological characteristics that determine the reference of 'milk.' 'Milk' refers to all kinds of mammalian excretions meant to nourish young (cf. Copp 2000). We do not think that 'milk' is nevertheless like 'water' in that it purports to rigidly refer to the essential internal nature of all milky stuff (only to fail for failure of singular reference). Why not? It seems silly to claim that some concepts just do purport to rigidly refer, and we have been lucky that the world has cooperated to deliver a singular reference for the rigidly referring terms, and multiple referents for non-rigid designators. The better explanation is that the rigid referrer 'water' only comes to rigidly refer after the discovery of a uniform internal nature to all watery stuff so that all our interests in water can be satisfied with information about H_2O. Whether there is a different concept after empirical supplementation depends on re-identification of concepts over time, though it seems that a precisification of ‘water’ in this manner would not yield a completely different concept.

Applied to ‘moral judgment,’ then, the thought is that empirical information about evolutionary history can come to *augment* and *precisify* the concept of moral judgment (as the information becomes widely appreciated), and can thereby narrow its extension to exclude reference to swamp judgments. This can be a gradual process through diachronic shifts in concept. While not a knock down refutation of swamp intuitions, considerations like these should give us pause when we decide how much weight to accord them.[19] Whether the original MJP or the revised version ultimately succeeds, both present improvements on the orthodox internalist and externalist thesis by respecting more intuitions on cases, and by offering a determinate conception of a society-wide background connection between moral judgment and motivation. However, some of the interesting aspects of MJP concern what we can say about actual amoralists and their malfunctions (Sect. 4), and for these claims we need not rely on the necessity claim.

The necessity claim to one side, we should take pains to avoid misunderstandings of purposive talk. Evolutionary explanations do not claim that moral judgments of the sort we have are necessary to perform the social function they in fact perform. Indeed, that function could be served in other ways. To take an analogous case, evolutionary explanations would not claim that no other phenotypes could serve the function that our hands actually serve. That is obviously false. In both cases, evolution purports to explain why our moral judgments or hands did in fact evolve given contingent facts about available phenotypes in our evolutionary environment and the functions performed by those phenotypes. So the proposed synthetic necessity claim is not that if a mental state M serves this function, then M is a moral judgment, but rather, if some mental state M does not have a certain social function as a purpose, then M is not a moral judgment.

Although I have thus far identified moral judgments by referring to their contents (e.g., keeping promises, not thieving, not harming others, helping others), it seems that other characteristics of first person moral judgments are ideally suited for their social motivational role, and so would be selected for. For example, these judgments would have to be motivationally weighty to override motivations for non-moral behaviors, at least in many cases. Otherwise the benefits of mutual cooperation would rest on less reliable foundations. However, it should be implicit that the motivation supplied by moral judgments is not in all cases sufficient to generate action. In other words, it is overridable. We can observe that the motivation to be moral changes and yields to other motivations when dealing with others who seem untrustworthy. And there are occasions when moral motivation is rightly outweighed by a concern for personal well-being. To illustrate, consider a case where you have promised to meet a colleague for lunch, and when the time comes you judge that you ought to keep your promise. Sadly, while walking to the

[19] This theory of conceptual change might help other etiologically minded philosophers in dealing with swamp intuitions. Alternatively, one could say that actual moral judgments and swamp moral judgments are different species of moral judgment, and the synthetic necessity claim applies to the species of actual judgments. For a defense of historical, biological kinds as natural kinds see Millikan (1996).

rendezvous point, you are hit by a car. On this occasion you would be more motivated to attend to your injuries rather than attend the meeting come what may, and rightly so. Overridable moral motivations like these are and were likely to be more fitness enhancing than non-overridable sufficient moral motivation.[20]

One thing we need from a theory of moral motivation is to preserve the fact that different people come to very different moral convictions, and each differing moral judgment tends to carry with it some motivational import. Yet from the above it might sound like we can only explain the motivational force of moral judgments that correspond to prototypical moral behaviors, like helping behaviors. It is important to note, however, that the above comments try to explain the existence of a motivational mechanism by appealing to the kinds of behaviors it historically helped generate, and once we discover that the purpose of moral judgments is to motivate prototypical cases of moral behavior, the machinery that evolved to do this can be co-opted by other practices that differ significantly from their proper function.[21] When this happens we should expect the co-opting normative judgments to motivate corresponding behaviors even if, historically, these behaviors were not the socially adaptive ones. That is, judgments that make use of the evolved machinery for making moral judgments will typically have corresponding motivations because the evolved machinery doesn't know any better.

Some concerns about the view will focus on its adaptationist underpinnings. The evidence suggests that some moral behaviors—including those that biologists call altruistic or cooperative—were selected because they made moral actors as a whole more fit than their non-moral competitors, and this increased fitness was generated in part by assortive interactions. Though different social groups could have developed slightly different ways of achieving these mutually beneficial outcomes, one might wonder whether this perspective makes moral behaviors look too monolithic. If moral behavior is merely fit behavior, then how do we distinguish moral from other kinds of behaviors that also contribute to fitness? In reality, our moral lives are very complex and one might wonder whether evolutionary modeling does justice to the complexity.

The MJP view does not attempt to boil down all of morality to a single function that it is meant to serve, though some philosophers appear to do just that. Arguably, Hobbes (1994 [1651]) thought that moral practices are just those things that we need

[20] See also Joyce (2006, p. 109). There might be particular cases where moral judgments are defective because they do not "win out" amongst competing motivations. The claim here is that insufficient moral motivation is not *necessarily* a defect. Also, an anonymous reviewer has noted that the purposive perspective might be extended to reveal that certain failings in other psychological states or activities, like willing, count as defects. The success of these extensions would have to be considered on a case-by-case basis.

[21] Haidt and Craig (2004) say something similar.

> Of course, it is possible to teach children to be cruel to certain classes of people, but how would adults accomplish such training? Most likely by exploiting other moral modules. Racism, for example, can be taught by invoking the purity module and triggering flashes of disgust at the 'dirtiness' of certain groups, or by invoking the reciprocity module and triggering flashes of anger at the cheating ways of a particular group (Hitler used both strategies against the Jews). In this way, cultures can create variable actual domains that are much broader than the universal proper domains for each module (p. 63).

to solve the problems we face in the state of nature, where the problems can be modeled by a prisoner's dilemma. Similarly, Gauthier (1986) has argued that moral norms are rationally pursued and adopted when self-interested individuals are faced with certain game-theoretic problems. By contrast, MJP is sympathetic to a plurality of moral practices that could each have a different purpose or purposes as complex responses to local environments, and influenced by a mixture of genetic and cultural selection.[22] If the analogy with the evolution of biological entities is any indication, the variety of purposes served by our moral practices could be as rich as the variety of functions served by our bodily organs.

4 Back to the amoralists

When we apply the MJP theory to amoralist cases we get intuitively right and more determinate results than mere reflection on cases. First, MJP permits the existence of some isolated amoralists, which nonetheless render genuine moral judgments. The purposive perspective can explain how some moral judgments with the right history—and so the right proper function—could be nonetheless be *defective* and fail to perform their proper function. Individuals who make moral judgments about what they ought to do, but fail to be motivated by them, render defective moral judgments (but moral judgments nonetheless). Though they render moral judgments, those judgments are not doing what they are *supposed* to be doing. This is no different than the discovery of defective left ventricles that are supposed to perform a certain function but fail to do so.

Orthodox externalism can also accommodate amoralists, but MJP does a better job at identifying and explaining the thought that the *moral judgments* of isolated amoralists are defective. Externalism can explain how amoralist *individuals* are defective, for they are morally defective. But orthodox externalism cannot capture a sense in which the amoralist moral judgment seems to be defective. MJP can. Just as the spark plug that fails to fire is defective insofar as it fails to do what it is supposed to do, on the MJP view moral judgments that fail to translate into appropriate motivational states are defective insofar as they fail to do what they are supposed to do. So in addition to explaining the moral failings of amoralist agents, we can identify and articulate a descriptive failing of moral judgments.

Second, as a synthetic necessity claim MJP has a rather refined view of the Amoralsville case and how it differs from isolated amoralists in our own community. Under MJP, whether or not a particular social group has moral practices depends on whether or not bits of thought, language and behavior were selected and propagated in the past because those practices elicited moral behavior. And moral judgments play a part in that system of moral practices. A community

[22] Compare those norms and attitudes that are appropriate within the realm of family and friends with the norms and attitudes that are appropriate within the realm of politics. In politics we do not believe it is appropriate to favor ourselves and those close to us, but in our private lives we do believe it is not only permissible, but also imperative to concern ourselves primarily with the well-being of those close to us. The purposive perspective can explain the variation by appealing to the different purposes that moral practices evolved in these two domains.

like ours, with only a few, isolated amoralists, would evidence community-wide moral practices and moral judgments. Isolated amoralists render moral judgments that do not fulfill their purpose, just as we find a few individuals with left ventricles that do not fulfill their purpose. Recall that Amoralsville residents generally behave in ways that respect moral norms, but they are never motivated by putative moral judgments. In fact, they learned about moral language from our community, and in Amoralsville moral categories are merely classificatory. Though most citizens purport to make judgments about their moral obligations, no citizen is thereby motivated. What gets them going is the fear of punishment and their own self-interest. Here it looks like Amoralsville did not develop a system of thought, language and behavior that facilitated mutually beneficial social interactions. As a result, they did not inherit moral judgments as part of those practices. Unlike the isolated amoralist, who has a mental state that counts as a broken moral judgment, Amoralsville residents do not even have broken moral judgments. Though the resident can descriptively pick out occasions of moral obligation, there is an essential motivational aspect of moral judgments that goes missing. In short, these mental states are not *supposed* to motivate in the requisite way, and so they do not count as genuine moral judgments.

Stepping back, we can see how background, external conditions might play some role in determining whether an individual has a particular mental state. If the mental state type M is essentially part of a developed social practice—as seems to be the case with moral judgments—then we have to look to the history of one's internal psychology and how it relates to the social practices in one's community to determine whether a particular token state m is of the type M. There probably is no bright-line level of motivation that determines whether a community is more like our own, where amoralists render moral judgments, or more like Amoralsville, where they do not. What is important is the history of interaction and the development of attitudes and other practices that enable cooperative social behaviors, which can occur in stages and degrees.

Third, the synthetic necessity claim aside, MJP is stronger than orthodox externalism insofar as the orthodox view relied upon statistical and nomological regularities without going so far as to say that actual moral judgments are supposed to motivate. In the actual world moral judgments motivate most of the time, but more than this we can say that they have motivation as a purpose. And as a synthetic necessity claim MJP offers a theory of how moral judgments are essentially action-guiding, not merely contingently action-guiding. To be sure, there is a sense in which it was entirely contingent whether moral practices and so moral judgments evolved to do the work that they in fact do. But the claim here is that, assuming that moral judgment did evolve to do this work, the connection between moral judgment and motivation is not entirely contingent. Given the way the world is, it is not possible to have certain community-wide failures of moral motivation. As indicated earlier, the success of this synthetic necessity claim should depend not only on intuitive evidence and how our intuitions might change as etiological information becomes more widely appreciated, but also on how well the view fares against other theories of moral judgments. On that second score it does rather well. It accommodates most intuitions and provides reasonable, determinate results about various cases of amoralism.

5 Conclusion

We now have two sources of support for a modest motivational internalism, one based on intuitions on various amoralist cases, and one based on the evolutionary history of moral practices. How do these views bear on other positions in metaethics? Regarding moral metaphysics, the present view fits in well with a naturalized ontology, for nothing here depends upon the existence of non-natural facts, or some genetic fallacy that improperly bridges an is-ought gap. That certain behaviors were selected for shows us that they have a purpose, but that kind of purpose does not translate into a reason for action, or any conclusion about what we ought to do in some normatively weighty sense any more than the fact that a car's ignition is supposed to start the engine gives us a reason to start the engine. This point is often overlooked because 'should do' can mean what we have most reason to do, or it can mean what we expect something to do, or it can mean what something would do if it functioned properly. MJP only invokes this third sense of 'should.' The present view is also consistent with ethical non-naturalism, though non-naturalists and realist naturalists alike must respect the fact that there is more to moral judgments than merely cognizing the ethical facts. There is a motivational component.

This brings us to moral semantics. Internalist theses are typically enlisted to support some version of non-cognitivism, for a widely held Humean view holds that cognitive judgments have no motivational force of themselves. Thus, if orthodox motivational judgment internalism were true and moral judgments were necessarily motivating, then they would be a kind of non-cognitive state. Under MJP, moral judgments do not necessarily motivate, but it is their purpose to motivate, and in favorable circumstances they do have some motivational force. One might look at those cases where moral judgments perform their purpose and motivate, and conclude that moral judgments are non-cognitive even on the purposive view. The basic inference is this: if ever motivational force, then non-cognitive state. To my mind, this is a mistake. It fails to respect the way in which isolated life-long amoralists still render moral judgments. Indeed, the view that emerges from reflection on various amoralist cases, and from the purposive view, inclines toward a besire theory where moral judgments play double duty as representing certain facts and motivating the appropriate behaviors.[23] Possible amoralists suggest that there is some cognitive element, and the fact that moral judgments are supposed to motivate, and often do, suggests a non-cognitive element (along with the bizarreness of attributing to agents some *de dicto* desire to do the right thing). Of course, the final answer concerning the structure of motivation should be left to empirical psychology. If further empirical work reveals that these roles (which are conceptually separable) are not realized by a single mental state token in individual psychologies we might choose to modify our conception of a moral judgment to more narrowly refer to one of these tokens, either the cognitive or the

[23] This term is due to Altham (1986). The most forceful arguments against besire theory depend upon the separability of cognitive and non-cognitive functional *roles* (cf. Smith 1994, Ch. 4), but these arguments fail to refute the claim that these separable roles are actually realized by one and the same mental state token in our moral judgments. For a detailed analysis of this claim, see Bedke (ms).

non-cognitive.[24] For now it is safe to say that defensible internalist theses do not do the work non-cognitivists have hoped for, and the possibility of certain kinds of amoralism does not do all the work that cognitivists have hoped for.

Acknowledgments I thank David Schmidtz and Mark Timmons for comments on an early draft of the paper. I also thank an anonymous reviewer for extensive and helpful comments on the submitted manuscript.

References

Altham, J. E. J. (1986). The legacy of emotivism. In: G. Macdonald & C. Wright (Eds.), *Fact science and morality* (pp. 275–288). Oxford: Basil Blackwell.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Bedke, M. (manuscript). A case for besires. Unpublished paper.

Brink, D. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.

Copp, D. (2000). Milk, honey, and the good lie on moral twin earth. *Synthese, 124*, 113–127.

Darwall, S. (1983). *Impartial reason*. Ithaca, New York: Cornell University Press.

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

Dreier, J. (1990). Internalism and speaker relativism. *Ethics, 101*, 6–26.

Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society, Sup. Vol.* 47, 187–208.

Gauthier, D. (1986). *Morals by agreement*. Oxford: University Press.

Gert, J., & Mele, A. (2005). Lenman on externalism and amoralism: An interplanetary exploration. *Philosophia, 32*, 275–283.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 181–218). Oxford University Press.

Haidt, J., & Joseph, C. (2004). Intuitive ethics. *Daedalus, Fall, 2004*, 55–66.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3). New York: Oxford University Press.

Hamilton, W. D. (1963). The evolution of altruistic behavior. *American Naturalist, 97*, 354–356.

Hobbes, T. 1994 [1651]. *Leviathan*. Edwin Curely (ed.), Hackett Publishing.

Hull, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics, 11*, 311–332.

Joyce, R. (2006). *The evolution of morality*. Cambridge, Mass: MIT Press.

Kitcher, P. (2006). Between fragile altruism and morality: Human evolution and the emergence of normative guidance. In G. Boniolo (Ed.), *Evolutionary ethics and contemporary biology* (pp. 159–177). Cambridge, Mass: Cambridge University Press.

Lenman, J. (1999). The externalist and the amoralist. *Philosophia, 27*, 441–457.

Mele, A. (1996). Moral cognitivism and listlessness. *Ethics, 104*, 727–753.

Mele, A. (2003). *Motivation and agency*. New York: Oxford University Press.

Millikan, R. (1984). *Language, thought, and other biological categories*. MIT Press.

Millikan, R. (1996). On swampkinds. *Mind and Language, 11*, 103–117.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York: Oxford University Press.

Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press.

Richarson, P., & Boyd, R. (2005). *Not by genes alone*. Chicago: University of Chicago Press.

Schafer-Landau, R. (2003). *Moral realism: A defense*. Oxford: Oxford University Press.

Skyrms, B. (1996). *Evolution of the social contract.* Cambridge University Press.

Skyrms, B. (2004). *Stag hunt and the evolution of social structure*. Cambridge University Press.

[24] Though this choice would not be forced upon us. After discovering the distinction between jadeite and nephrite, we continue to call them both 'jade.'

Smith, M. (1982). *Evolution and the theory of games*. New York: Cambridge University Press.

Smith, M. (1994). *The moral problem*. Oxford: Basil Blackwell.

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge: Harvard University Press.

Timmons, M. (1999). *Morality without foundations: A defense of ethical contextualism*. New York: Oxford University Press.

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