Moral realism and teleosemantics
Richard Joyce

In a recent article, William F. Harms (2000) argues in a novel way for a form of moral realism. He does not actually argue that moral realism is true, but rather that *if* morality is the product of natural selection—as many people (including myself) think—then moral realism may be secured. The thesis that allows the bridge is a certain conception of truth conditions (for which he is indebted in particular to Ruth Millikan (1984, 1993)[1]): teleosemantics. Thus—though Harms is less explicit about this—his argument is really doubly conditional: moral realism depends not only on the truth of an empirical hypothesis concerning the evolution of morality, but also on the success a substantive and controversial philosophical program. In this short paper I am not going to launch an attack on teleosemantics, but we should be aware throughout that it is hardly a thesis beyond question, and there are many respectable detractors.[2] For now, however, let us grant both conditions, and see whether a route to moral realism opens up. I will argue that it does not.

Harms' argument is disarmingly simple. According to Millikan (*et al.*) certain signals—such as bee dances, vervet monkey warning cries, etc.—may be considered true just in case they fulfill the function they are supposed to fulfill: that function assigned by natural selection. Putting things crudely: such signals are not useful because they are true, but are true because they are useful. If morality is an adaptation, then a token moral judgment is true just in case it is fulfilling its evolutionary function.

This reveals that the hypothesis upon which Harms' argument depends is not merely that "human morality is the product of natural selection", but that "human morality *as a system of signals* is the product of natural selection". That the former does not imply the latter can be seen by considering the hypothesis that what the pressures of natural selection favored (in humans) was a range of moral *emotions*: disgust at unprovoked violence, aggression towards defectors, guilt concerning temptations to cheat one's fellows, and so on. Imprecise as that is, it would suffice for us to speak of "morality being an adaptation", but it would not follow that what has evolved is the right kind of thing (i.e., a signal) to be considered *true*, any more than a properly functioning human heart or healthy fingernails might be considered true.

Let us, then, allow for the sake of argument that human morality has evolved as a system of signals. It might help to have a comparison example: there is some evidence that humans have evolved a natural aversion to snakes.[3] We need to suppose not merely that this aversion manifests itself as an emotion ("repulsion"), but also that natural selection has exerted its pressures beyond the emotional sphere, selecting for particular verbal responses as ways of expressing these emotions: people naturally utter "Yech!" in the presence of a snake. Yet even this will be insufficient for the

---

[1] See also Dretske (1995), Godfrey-Smith (1996).
[2] See Fodor (2000), Botterill and Carruthers (1999), chapter 7.
[3] See, for example, Ohman (1979), Mineka *et al.* (1984).

conclusion that when humans say "Yech!" at snakes they are saying something true. "Yech!" just doesn't have the right sort of grammar. Can we use it as a premise in an argument? As the antecedent of a conditional?[4] Can we even sensibly respond to someone's uttering it with "That's true"? The evolutionary hypothesis evidently must be made even more precise: that human morality as a system of signals *that function to indicate that something is the case* (i.e., as a system of representations) is the product of natural selection. Returning to the snake analogy, we must assume that the tendency to say (or think[5]) something like "There is a dangerous snake!" in the presence of snakes is an evolved trait.

So specified, the evolutionary hypothesis begins to lose some of its attractions. To see this, we need to ask a difficult question: On the assumption that moral discourse has evolved, what has it evolved *for*? I cannot here do this question justice, so a sketch must suffice. Many people find it illuminating to think of the evolution of morality in terms of the Prisoner's Dilemma (see, for example, Axelrod (1984), Trivers (1971)). Morality is like an in-built preference for cooperation, allowing agents to enter into stable cooperative ventures. The other side of the coin is that morality involves an antipathy towards defection. The temptation to defect on a deal makes a moral person feel guilty; moreover it licenses an aggressive response to any perceived cheating (thus regulating a tit-for-tat strategy—or something like it—in iterated Prisoner's Dilemma games). Let's not complicate things beyond that. Consider, then, a piece of moral language: upon perceiving Jack defect on a deal, Susan says "He ought not do that; it's morally wrong". The evolutionary function of this language, we might say, is not to describe the world, but to license punishment of Jack, or to strengthen the bonds that Susan enjoys with other cooperators. If this is the case, then the utterance is like "Yech!" for snakes: not the right kind of signal to count as truth-evaluable. Unlike "Yech!", of course, this moral utterance does come in indicative form, and we can imagine Susan using it as an antecedent, etc. But still, it is not to be simply assumed that the *evolutionary* function of a sentential signal— even if that signal comes in the indicative mood—is to indicate that something is the case.

Later we'll come back to the possibility that the function of moral utterances is something other than indicating that something is the case, but for now we'll put this aside and allow that they do naturally function in this manner. The main concern that I wish to address is that natural selection may nevertheless have favored a systematically false representation. Let's stick with the snake example…

Even supposing that a representation like "There is a dangerous snake" (firing in the presence of snakes) is an evolutionary trait, it would be silly to see it as just a noise we make in the presence of snakes. Unlike a vervet monkey's snake signal, the human utterance has internal structure: a subject and a predicate that can be applied to other subjects: "There is a dangerous crocodile", "There is a dangerous overloaded power socket", etc. Indeed, we can say general things about what "It is dangerous" *means*: "It is likely to cause harm". This observation is not tantamount to a rejection of teleosemantics, but a demand that it achieve a degree of sophistication.

---

[4] See Geach (1960).

[5] Harms is unclear whether he takes the product of evolution to be a public signal or a private thought. In order not to second guess him, I too am leaving it open.

Teleosemantics is not going to be a very exciting theory if it can fix the meaning only of words and expressions useful in the ancestral environment. Any interesting version must account for the introduction of new linguistic terms via the old, and must thus recognize the complex links that hold among pieces of language. For example, if a teleosemantic account can fix the meaning (or, if you prefer, provide the truth conditions) of the sentences "It is likely", "X causes Y" and "It is harmful", then it can fix the meaning (provide the truth conditions) of "It is dangerous". This kind of inferential meaning-fixing must trump any competing analysis according to which its meaning is established by environmental stimuli. This is evident from the fact that even if our first reaction is to think of the grass snake as dangerous, reflection on the matter, noting the absence of any menacing equipment on the animal, will lead us to revise our claim.[6]

As with snake language, so with moral discourse. Public moral judgments are not merely noises we make in response to environmental stimuli—they must get their meaning fixed by reference to other language. In other words, there are central claims that we make concerning them: *a priori* platitudes that surround our moral language, such that a failure to endorse such a claim will be taken as an indication of linguistic incompetence. Some have argued, for example, that it is an *a priori* platitude about morality that it is necessarily the case that if a practically rational person judges some available action ϕ to be morally required, then he will have some (defeasible) motivation in favor of ϕing.[7] Another platitude (more plausible, in my opinion) is that if a person is morally required to ϕ, then she ought to ϕ regardless of whether doing so promises to satisfy her desires.[8] One might argue that there is a tight connection between "S ought to ϕ" and "S has a reason to ϕ".[9] Another truism is surely that morally required actions are of great (perhaps overriding) importance. Michael Smith (1994, pp. 39-41) supplies a whole raft of moral platitudes: "When A says that ϕ-ing is right, and B says that ϕ-ing is not right, then at most one of A and B is correct", "Acts with the same ordinary everyday non-moral features must have the same moral features as well", "Right acts are often concerned to promote or sustain or contribute in some way to human flourishing", and so on.

Putting together a catalogue of such platitudes is not an easy job, and it may well be that some of those above should not even be on the list. But the point is that there certainly *is* a rich set of convictions surrounding our moral concepts and judgments, and this set provides a constraint on when an utterance of the form "X ought not do that" is true and when it is false. If we say of some action, ϕ, that Jack morally ought not perform it, but it turns out that ϕing falls significantly short of satisfying the platitudes about prohibitions that we endorse (when we reflect carefully on the matter), then we have said something false. This will remain so even if the utterance is fulfilling its evolutionary function (assuming it has one).

But could our moral discourse be *systematically* false—enough to undermine moral realism? Maybe. A certain type of moral skeptic (of which John Mackie is a

---

[6] Exactly how this "trumping" works I don't pretend to know. All I am claiming is that the teleosemanticist will need to recognize and account for it.

[7] See, for example, Smith (1994), Korsgaard (1986).

[8] See, for example, Kant (1983).

[9] See, for example, Mackie (1977), p. 77.

prime example) will hold that the platitudes surrounding moral discourse are such that *nothing* satisfies our moral judgments. Suppose—just to have an example—that we firmly endorse (i) that if S ought to φ then S has a reason to φ, and (ii) that if S morally ought to φ, then S ought to φ regardless of S's desires; then we could conclude that if S morally ought to φ, S may have a reason to φ regardless of S's desires. But then we might see that according to our best theory of reasons—perhaps even that provided by a teleosemantic analysis—there are *never* any reasons of this sort. The conclusion would be that no sentence of the form "S morally ought to φ" is true.[10] I am not claiming that it *is* this way with moral utterances, only that it could be. That would suffice to show that the evolutionary hypothesis (which has now been refined), plus teleosemantics, does not lead straight to moral realism.

To show that something has gone wrong with Harms' argument, it may help to consider another analogous example. There has been some evidence put forward (not very rigorously, as far as I know) that humans have evolved to be religious creatures.[11] We are naturally disposed to believe in God. Let us speculate that this not entirely implausible hypothesis is true (it doesn't matter to my point if it's not). Perhaps reproductive fitness was well served if humans believed there to be some all-powerful authoritative, loving being, observing human affairs and doling out rewards and punishments accordingly. (Perhaps the usefulness could be understood again in terms of regulating cooperative behavior, thus avoiding Prisoner's Dilemma scenarios.) Suppose that various basic religious utterances have been selected for: "God will punish that kind of behavior", "God made the world", "God exists", etc. As far as I can see, according to Harms such utterances must automatically be true. But that just *can't* be correct. The fact that such beliefs might have helped our ancestors make babies does not entail that an omniscient, omnipotent being is overseeing the world!

In light of these kinds of considerations, the teleosemantic moral realist may try a different tack. Despite in places confidently promising that with teleosemantics will come moral realism—that is, an account of how moral judgments might be *true* (assuming the evolutionary hypothesis)—when it comes to the crunch Harms is willing to back off. He allows that perhaps moral signals do *not* function to indicate that something is the case, but rather like commands—and commands, he is willing to concede, may not have the appropriate structure to be considered truth-evaluable.

> If you prefer to reserve the term "truth" for the correspondence of propositions to the world, I won't quibble with the stipulation. What matters is that there are objective conditions for the correctness of the issuance of a particular signal at a particular time and place, and that they rely on objective semantic maps (p. 704).

But this is more than a terminological quibble; it's a shift in the rules of the game: teleosemantics was supposed to provide moral realism, but suddenly realism doesn't require truth, it merely requires "objective conditions for correctness of issuance". Harms needs to be more careful with his promises, because for many theorists

---

[10] To see an attempt to nail this argument down in detail, see Joyce (2001).

[11] See, for example, Boyer (1994), Ramachandran and Blakeslee (1998), D'Aquili and Newberg (1999).

"objective conditions for correctness" falls short of the criteria for a *realistic* construal of a discourse. But for the sake of argument we'll allow Harms his broad understanding of realism (perhaps we should call it "quasi-realism"), and see whether he has paved the way even for that.[12]

Let's go back again to the snake-signal adaptation, but this time construing it as an imperative—say, "Look out for that snake!" Now it is not entirely clear to me what is supposed to count as the evolutionary function of such a command. Presumably it cannot be the bringing about that one's addressee avoids the snake, or else we'd have to conclude that if the addressee misunderstands or mishears, and hence is bitten, then the speaker's command was incorrect. (In which case, correct moral condemnation of, say, genocidal torturers, would then require that the guilty *heed* the denunciation!) Perhaps instead the idea is that a command is correct if uttered in a situation in which the state of affairs it is supposed to bring about (that is, the kind of state of affairs the bringing about of which provided improved reproductive fitness for ancestors) is not obtaining. In this case, when there is a snake at a person's feet, but he is not jumping away, then the imperative "Look out for that snake!" is correct (irrespective of whether he actually jumps). This sounds more palatable, but there are still good reasons for doubting that this yields realism.

Consider again the example of an evolved religious disposition. Perhaps whatever religious adaptations we have involve not just indicative language, but imperatival too. We are naturally disposed to say to the wrongdoer "Repent thy sins!" or "Praise the Lord". According to Harms, there will be standards of correctness for such utterances. Fair enough—we'll allow him whatever account he wishes. But the crucial question is: Does it provide the basis of theistic realism? Should such considerations upset the theistic skeptic (i.e., the atheist)? No and no. All it requires of the skeptic is that she accept that such religious language has been systematically useful, in such a way that its employment enhanced reproductive fitness. But the theistic skeptic can happily admit this; indeed, she is quite attracted to the idea that religion is nothing but a useful fiction.

The mere fact that we can make *sense* of the idea of a "useful fiction" shows that teleosemantics is not going to automatically secure realism. If what is *ex hypothesi* a fiction can be "useful", then it can, in principle, be favored by natural selection. But something is evidently wrong with the idea that as soon as natural selection gets involved the fiction immediately becomes objectively true. This goes for imperatival systems as well as propositional ones. Can we make sense of a seriously flawed command? Of course. If I say "Shut the window" when it is already shut, then my utterance is badly disoriented. (In so far as an imperative has representational implications, then such an imperative misrepresents.) But can we make sense of its being useful for a person on occasions to make such a mistake—whether intentionally

---

[12] Here I'm putting aside a crowd of possible objections. For example, there is an awkward question of moral judgments in the past tense, such as "Hitler ought not to have exterminated the Jews". Who exactly are we commanding? And what would count as a successful command in such circumstances? Even in the present tense, we often make moral judgments seemingly as a way of expressing our allegiance to a normative system, rather than trying to alter anybody's behavior or bring about a change in the world. For example, I will judge that stealing is wrong even if nobody is stealing, and nobody is even tempted to steal. I might say it in response to a fictional story about stealing, or after imagining the mere possibility of stealing.

or otherwise? Again, yes; we might have to tell a bit of an odd tale, but it's no great stretch of a philosopher's imagination. Commands that cannot be carried out may nevertheless have an effect on a person's behavior, and such effects may be useful. And if we can imagine its being useful on some occasions, then we can imagine its being systematically useful. We can imagine a situation in which the benefit provided (however slight) by uttering the flawed imperative has a positive impact (however slight) on an individual's reproductive fitness. And if we can imagine that agent's offspring being similarly situated, then we can imagine natural selection favoring the employment of this *mistaken* utterance.[13]

It is possible, then, that even granting that moral discourse is a system of naturally selected commands, the discourse is faulty. ("Faulty" sounds vague, but the metaethical vocabulary doesn't have an established term—analogous to "false" for statements—for criticizing commands. Usually the thesis that morality is a system of commands is itself considered a form of moral anti-realism.) Of course, the fault isn't going to be the simple one of saying "You ought to ϕ" when the addressee has already ϕed. The fault may instead be that the commands are made within a framework of well-entrenched beliefs and presuppositions that render the commands defective. What I have in mind is again the cluster of established platitudes at the heart of our moral discourse, gestured at earlier. A moral command like "Don't kill" will bear important relations to statements like "Killers ought to feel guilty", "Killers deserve to be punished", "If one really accepted that it's morally wrong for to kill, one would be motivated to refrain", "Killers are being irrational" (perhaps), etc. The legitimacy of the command "Don't kill" depends in large part on the legitimacy of these other central claims. And as we saw with moral representational language, if a substantial number of these statements turn out to be false, then the whole moral edifice is what the moral skeptic has always claimed it to be: an erroneous—though perhaps useful—way of speaking. (Let me emphasize that here I do not purport to present a cogent argument; my point is to sketch how natural selection *could* favor a normative discourse of commands that is systematically flawed.)

I accept the evolutionary hypothesis that the "moral faculties" of humans have been produced and worked upon by natural selection, hence I agree that we should seek to understand evolutionary pressures if we want to explain where morality comes from. I am much more tentative in accepting that moral language has been selected for. I have in this paper entertained the possibility that quite complex linguistic types might count as items that have been naturally selected, though as a matter of fact, I think it extremely unlikely that such signals might have been selected *as units*. Teleosemantics for human language is plausible only when it begins with the selection of the small and simple units, building the complex upon those foundations. Even if a range of simple moral signals are part of humans' evolutionary tool kit, it's an enormous jump to thinking that the rich and complex moral discourse that we now

---

[13] The teleosemanticist may complain that I'm begging the question, by fixing the meaning of "Shut the window" independently of, and prior to, the impact of natural selection. But we can grant that teleosemantics has fully fixed the meaning of this sentence for time t, while imagining that the pressures of natural selection are ongoing, such that at some point after t it becomes reproductively useful to employ the command in a faulty manner. The teleosemanticist may insist that this amounts to the meaning of the command changing. But why must this be?—the meaning of "shut", "the" and "window" may remain stable throughout.

employ might find vindication in fulfilling the same insular communal ends as did our ancestors' value judgments.

In order to see how eccentric this view really is, consider the epistemological question of how we are to discover the moral "truth". Harms is willing to concede that according to his view there are no guarantees of moral knowledge, but at least there is a truth "out there", depending on objective facts about our adaptive history. Suppose that according to our best theorizing concerning the archeological evidence, we decide that the social conditions through the relevant period of hominid development were thus and so, implying that morality evolved to serve some particular function. Accordingly, we could provisionally determine the truth conditions of contemporary moral claims: "Third trimester abortions are sometimes permissible", "Canceling third world debt is morally desirable but not obligatory", "The Elgin Marbles ought to be returned to Greece", and so on. But then imagine that further evidence from paleontology and prehistoric campsites unexpectedly turns up, suggesting radically different conclusions concerning the social arrangements of the relevant hominids, and thus leading us to revise our theory about the evolutionary function of morality. This could, according to teleosemantic moral realism, demand a revision of our contemporary moral views: in principle, what we had thought was probably true of, say, the cancellation of third world debt, we would now see is probably false. But it is surely an astonishing theory that holds that the complex moral disputes that face the modern world might, even in principle, be settled by digging in the African soil.[14]

References

Axelrod, R.: 1984, *The Evolution of Cooperation*, Basic Books, New York.

Botterill, G. and Carruthers, P.: 1999, *The Philosophy of Psychology*, Cambridge University Press.

Boyer, P.: 1994, *The Naturalness of Religious Ideas: A Cognitive Theory of Religion*, University of California Press, Berkeley, CA.

D'Aquili, E. and Newberg, A.: 1999, *The Mystical Mind: Probing the Biology of Religious Experience*, Fortress Press, Minneapolis.

Dretske, F.: 1995, *Naturalizing the Mind*, MIT Press, Cambridge, MA.

Fodor, J.: 2000, *The Mind Doesn't Work That Way*, MIT Press, London.

Geach, P.: 1960, 'Ascriptivism,' *Philosophical Review* 69.

Godfrey-Smith, P.: 1996, *Complexity and the Function of Mind in Nature*, Cambridge University Press.

Harms, W.F.: 2000, 'Adaptation and Moral Realism', *Biology and Philosophy* 15.

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

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

Korsgaard, C.: 1986, 'Skepticism About Practical Reason', *Journal of Philosophy* 83.

Mackie, J.: 1977, *Ethics: Inventing Right and Wrong*, Penguin Books, New York.

Millikan, R.G.: 1984, *Language, Thought, and Other Biological Categories: New Foundations for Realism*, A Bradford Book, MIT Press, Cambridge, MA.

Millikan, R.G.: 1993, *White Queen Psychology and Other Essays for Alice*, A Bradford Book, MIT Press, Cambridge, MA.

Mineka, S., Davidson, M., Cook, M., Keir, R.: 1984, 'Observational conditioning of snake fear in rhesus monkeys', *Journal of Abnormal Psychology* 93.

Ohman, A.: 1979, 'Fear relevance, autonomic conditioning, and phobias: a laboratory model', in Sjöden, D.O., Bates, S., Dockens, W.S., (eds.), *Trends in Behavior Therapy*, Academic Press, New York.

Ramachandran, V.S. and Blakeslee, S.: 1998, *Phantoms in the Brain*, Quill, New York.

Smith, M.: 1994, *The Moral Problem*, Blackwell, Oxford.

Trivers, R.L.: 1971, 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology* 46.